THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXXII April, 1941 Number 4

STUDIES OF HIGHER-LEVEL READING ABILITIES

FRANCIS P. ROBINSON AND PRUDENCE HALL
Ohio State University

I. READING ABILITIES FOR DIFFERENT SUBJECT MATERIALS

It is probably more correct to speak of a person's "reading abilities" than of his "reading ability." Such a concept must obviously influence the approach to the measurement of reading performance as well as to the teaching of reading.

It has been known for some time that reading ability in one field is not very predictive of reading abilities in other areas. For instance, Pressey and Pressey[1] long ago showed that different types of reading material (prose, poetry, and science) obtained by reclassifying selections used in standardized reading tests are not highly related. Also the low intercorrelations[2] between published reading tests indicate that they do not even measure comparable composites of reading abilities. From the point-of-view of psychological analysis, the main shortcoming of these, and of similar studies made later, is that the selections used vary greatly in style, format, testing procedure, etc., as well as in content. Also, the use of selections, which are made up of short paragraphs on different subject-matters and which are constantly interrupted for answering questions, are unlike the normal reading situation. These uncontrolled variables tend to obscure the relationships present. The present experiment introduces certain refinements in technique to obviate these difficulties. The selections, differing in subject-matter, are taken from a common editorial source, continuous narrative is used to resemble normal reading, the tests are long enough to produce quite reliable measures, and scores for rate and compre-

[1] Pressey, L. C. and S. L.: "A Critical Study of the Concept of Silent Reading Ability." *J. Educ. Psychol.*, Vol. XII, 1921, pp. 25–31.

[2] Gates, A. I.: "An Experimental and Statistical Study of Reading and Reading Tests." *J. Educ. Psychol.*, Vol. XII, 1921, pp. 303–314, 378–391, 445–464.

hension are separated by having the questions follow the timed reading.

This study deals with the relation of reading performance in the fields of art, fiction, geology and history. The program was carried on with college students where higher-level reading skills would be likely to appear; and, as will be seen later, an attempt is made to go beyond the initial intercorrelations to a further analysis of the reading abilities measured.

METHOD

Subjects.—Two hundred five students in eight sections of a required course in educational psychology in the College of Education at Ohio State University acted as subjects. This enrollment consisted largely of freshmen and sophomores, although there were a few juniors and seniors and one graduate student in the group.

Materials.—Five reading rate and comprehension accuracy tests were constructed; four different fields (art, geology, fiction, and history) were represented, and there were two selections in one field (Canadian and Russian history). All selections except fiction were taken from *Compton's Pictured Encyclopedia;* the fiction selection was taken from *Collier's Magazine.*[1] The texts were from three thousand to forty-five hundred words long and represented a complete article or stopped a little short of the end at a natural break.

Each test packet contained a direction sheet with spaces to record personal data and rate scores, the text of material to be read, and a set of questions on the selection. An attempt was made to frame questions in terms of the main meanings developed by each headed section rather than in terms of unimportant detail; questions involving common knowledge, however, were avoided. The order of questions on various phases of the reading followed the order of the development of these ideas in the text. At the left of each question was the number appearing beside the line in the text which the subject must have read in order to know the answer; each subject answered only questions covered by the material which he had been able to read within the time allowed.

---

[1] *Compton's Pictured Encyclopedia* Vol. C, pp. 58–64; Vol. F, pp. 36–42; Vol. GH, pp. 39–45; and Vol. QR, pp. 182–190. Chicago: F. E. Compton and Co., 1939. Day, Edgar: "First and Forever," *Colliers*, Vol. CII, No. 16, October 15, 1938, pp. 20–21, 38, 42.

Administration of Tests.—The order of giving the tests was rotated in each class so that any one test was given to one class on the first day and to another on the second day, or at the first of the hour in one class and at the last of the hour in another.

Four sections ($N = 93$) took four tests (Canadian history, art, geology, and fiction), while four other sections ($N = 112$) took only two tests (Canadian history and Russian history). Time did not allow each group to take all five selections and part of the second group was used for another experiment (see Part III).

Rate scores were in terms of the word count at the left-hand margin at the end of three, six and nine minutes. Due to very rapid reading, twelve rate scores (two per cent of five hundred ninety-six scores) for the final three minutes had to be obtained by extrapolation from the scores for the first two intervals.

Comprehension was scored as comprehension accuracy, *e.g.*, the per cent right of the questions asked on the text up to the line which a given student was able to read by the end of nine minutes.

RESULTS

The intercorrelations between the reading rate scores for Group I ($N = 93$) on the tests for art, fiction, geology and Canadian history are presented in Table I. The intercorrelations between compre-

TABLE I.—INTERCORRELATIONS AMONG RATE SCORES ON ART, GEOLOGY, FICTION, CANADIAN HISTORY, AND RUSSIAN HISTORY TESTS[1]

Test	Art	Geology	Fiction	Canadian	Russian
Art	.....	.94 ± .008	.74 ± .030	.82 ± .023	
Geology	.82 ± .023	.....	.59 ± .046	.74 ± .03	
Fiction	.66 ± .039	.54 ± .049	.....	.62 ± .042	
Canadian history	.71 ± .034	.67 ± .039	.58 ± .046	.....	.98 ± .002 ($N = 112$)
Russian history	.....	.....	.....	.85 ± .018	

hension accuracy scores on these selections for the same group are presented in Table II. The correlations between scores for Group II ($N = 112$) on the Russian history and Canadian history tests are included with the others in Tables I and II in order to provide a measure of consistency on two disparate topics in a given field. Both Groups I and II were selected at random from the same required course and their results proved to be comparable (Canadian history

was read by both groups: Ave. rate = 2168 and 2245, SD = 443 and 468; and ave. comp. accur. = 55.6 per cent and 55.7 per cent, SD 17.1 per cent and 16.4 per cent); the intercorrelations are, therefore, probably comparable.

TABLE II.—INTERCORRELATIONS AMONG COMPREHENSION ACCURACY SCORES ON ART, GEOLOGY, FICTION, CANADIAN HISTORY, AND RUSSIAN HISTORY TESTS[1]

Test	Art	Geology	Fiction	Canadian	Russian
Art	.....	.48 ± .054	.79 ± .026	.44 ± .056	
Geology	.25 ± .066	.....	.31 ± .063	.63 ± .043	
Fiction	.37 ± .060	.16 ± .068	.....	.17 ± .067	
Canadian history	.24 ± .066	.38 ± .060	.09 ± .069	.....	.96 ± .004 (N = 112)
Russian history	.....	.....	.....	.49 ± .049	

[1] Attenuated correlations are in right corner of table and unattenuated correlations are in left corner.

It will be observed that varying degrees of relationship are shown between scores on the rate tests for art, geology, fiction, and history. The attenuated correlation between scores on the two history selections (.984) is higher than the other correlations. The fact that such a high agreement was shown between scores on history selections dealing with two disparate topics strongly suggests that history reading represents a rather definite skill. Intercorrelations between scores on the fiction test and those on the other tests tend to show the least relationship, but in every case there is indication that reading rate for one subject field should be used with much reservation if reading rate in other fields is to be predicted.

Intercorrelations between the comprehension accuracy scores also show variability. The fact that these correlations (corrected for attenuation) are much lower than those among the rate scores indicates that there is less consistency in comprehension accuracy than in reading rate for different kinds of material and that it is particularly dangerous to predict comprehension ability from one subject field to another. The attenuated correlation between scores on the two history comprehension accuracy tests, however, shows a much higher relationship (.959) than do the other correlations, and again indicates a basic reading skill for history.

A further analysis of these divergent reading skills is afforded by a comparison of the rates of reading for each three minute period for each

selection. (See Fig. 1.) It will be noted that the average rate scores for each of the selections tend to be more comparable during the first three-minute period than during the later intervals. The great drop in rate during the middle three minutes of the geology test seems to be due to the presence of a table at a point in the text where most of the subjects were reading during the second interval.

The greater similarity of the rates during the first three minutes as compared to the other three-minute intervals is interesting. Does it mean that each individual tends to start each selection in a somewhat

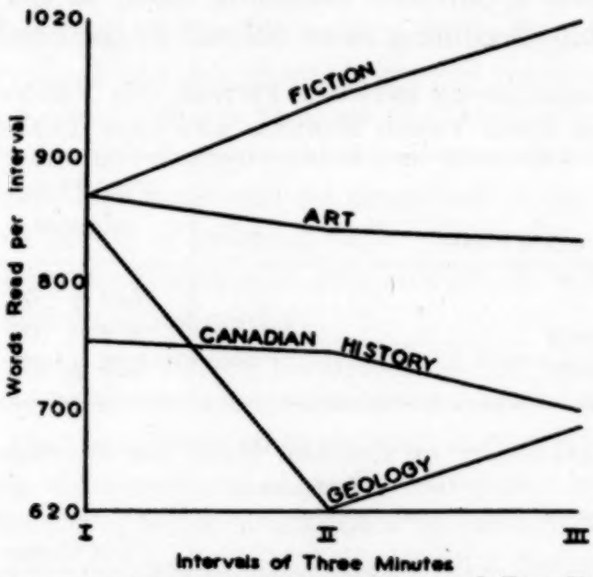

FIG. 1.—Average rate scores on art, geology, fiction, and Canadian history tests for each three-minute interval.

similar manner and then shifts to a "better" reading method for that subject-matter? Or does it mean that because the reader starts all of these selections with a "fresh mind" unburdened by problems of comprehension he is able, in the case of non-fiction, to comprehend at a fair rate until the number of ideas being assimilated demands a slower pace; and, in the case of fiction, the early setting of the story allows the reader to facilitate his comprehension of the later simple ideas of the story? Each hypothesis leads to different teaching and measurement techniques, so further analysis seems pertinent.

As a test of these hypotheses, correlations were computed for fiction *vs.* non-fiction—the most distinct difference in this graph. These uncorrected correlations are presented in Tables III and IV. It will be seen that the relationship between rates for these two types of material during the third three minutes of reading is not reliably

different from the relationship during the first three minutes. Further, in comparing rates for various three-minute periods on the same subject-matter, the first and second three minutes show as high a relationship as between the second and third three minutes. The nature of reading different subject-matters seems to be dissimilar, or as similar, during the first three-minute as during the third three-minute interval. On the basis of these results, it may be said, at least tentatively, that the divergence in rate between fiction and non-fiction seems to be due to differences in the speed of assimilating the ideas expressed —this assimilation apparently becoming easier as the fiction story is read and gradually becoming more difficult in the academic subjects.

TABLE III.—CORRELATIONS BETWEEN FICTION AND NON-FICTION RATE OF READING FOR FIRST THREE MINUTES AND LAST THREE MINUTES

	First three minutes	Third three minutes
Fiction *versus* art	.543 ± .05	.483 ± .05
Fiction *versus* geology	.572 ± .05	.441 ± .06
Fiction *versus* history	.535 ± .05	.482 ± .05

TABLE IV.—RELIABILITY OF READING RATE FOR DIFFERENT SUCCESSIVE INTERVALS

	First three minutes *versus* second three	Second three minutes *versus* third three
Art	.636 ± .04	.656 ± .04
Fiction	.855 ± .02	.836 ± .02
Geology	.744 ± .03	.735 ± .03
History	.780 ± .03	.742 ± .03

The fact that reading performance for a given field maintains its relation to other fields from the first is important in light of the present tendency in standardized rate tests to test over short intervals of time. The results in Fig. 1 do indicate, however, that rates over short periods of time are not indicative of production over longer intervals of time.

CONCLUSIONS

(1) Reading in different subject fields is not highly related even though these selections are written under one editorship and tested

by similar techniques. On the other hand, reading in different topics (Russia and Canada) in a given field (history) is quite consistent. It seems evident that reading tests for predicting reading ability in an academic subject should use selections from that field.

(2) This difference in reading performance on various subject-matters apparently shows up from the first. Intercorrelations for the first three minutes of reading are about the same as those during the third three minutes.

II. SOME QUALITATIVE DIFFERENCES BETWEEN GOOD AND POOR READERS

The preceding analysis has shown that in reading different subject-matters an individual will demonstrate several reading abilities. The purpose of this next analysis is to show that individual differences for a given selection are not simply a matter of quantitative differences in speed or accuracy, but may well be described in part as qualitative differences or as levels of reading skill. Basically, this will be an analysis of reading performance at different levels within the group—poor, average, and superior readers.

In order to make the results from each of these subsections of the total range more reliable, other cases were added to those who had participated in Part I. Other class sections in educational psychology were tested—ten class sections in a How-to-Study course were used and several groups in a WPA recreational leadership training program participated. Thus the group which had read a given selection varied in size from two hundred eighty-five to three hundred twenty-two.

In Part I the lack of high intercorrelations, except in the same subject field, and the fanning out of reading rate with successive intervals of time when fiction and non-fiction were read, showed what might be termed "flexibility" in reading adjustment from selection to selection. A comparison of good and poor readers, however, shows that this flexibility is not universal. This can be most easily demonstrated by noting the degree of flexibility of the highest fifth and the lowest fifth of each group in reading the various reading tests. The average (and its probable error) number of words read in nine minutes for each selection for each of these extreme groups is presented in Table V.

The poor readers make relatively little differentiation in their speed of reading selections which vary markedly in difficulty. The maximum difference for poor readers (one hundred forty-four words in nine minutes) is between a simple fiction story and a dry, factual account

of geological eras. Of six possible intercomparisons of rate for poor readers on the different subject-matters, only one difference is statistically significant. On the other hand, good readers do adjust to this shift in difficulty. The maximum difference in rate for good readers is eight hundred fifteen words in nine minutes and all of the differences in the six intercomparisons are statistically significant. It should be

TABLE V.—FLEXIBILITY IN THE RATE ADJUSTMENT TO SUBJECT DIFFICULTY BY THE FASTEST AND SLOWEST READERS

	N. each subgroup	Mean rate lowest fifth	Mean rate highest fifth
Art	62	1496 ± 19	3014 ± 32
Canada History	67	1500 ± 16	2804 ± 31
Fiction	75	1574 ± 17	3463 ± 33
Geology	57	1430 ± 14	2648 ± 23

noted also that this gamut of selections must actually represent a greater functional range for poor readers than for good readers.

Figure 1 showed that during successive three-minute intervals rate tended to decrease with non-fiction material and to increase with fiction material. It can be demonstrated, however, that poor readers do not show this flexibility during the reading of these selections whereas better readers do.

Inspection of different levels within the total range of speed of reading showed changes in patterns over the successive three-minute intervals. The three characteristic levels, which seemed to exist in the art and history selections, are shown in Table VI. The two levels found in fiction reading are shown in Table VII. The geology selection had a table in the text and also a marked shift in style during the article, so it was not possible to compare good and poor readers over successive intervals on comparable material.

In reading art and history, slow, average and fast readers did not all follow the declining rate pattern of the whole group (as in Fig. 1). These differences in pattern at the various rate levels are indicated by the differences between the rates for the first and the third three-minute intervals. See Table VI. The Critical Ratios given in this table are in terms of the standard error of the means for these periods.

Fast readers (about the upper fifteen per cent) on the art and history selections seem to be able to maintain their speed of assimilation or possibly to increase their rate with passage of time. There is no such

decrement in rate as is shown for the whole group in Fig. 1. "Average" readers, on the other hand, find that the process of assimilation for such non-fiction articles as these becomes somewhat "bogged down" (statistically significant difference) as they get into the article. The number of ideas to be assimilated keeps piling up so that they must read more slowly. If such average readers have difficulty with comprehension and so must adjust their rate, one might expect poor readers

TABLE VI.—AVERAGE CHANGE IN RATE ON NON-FICTION DURING THIRD THREE-MINUTE INTERVAL COMPARED TO FIRST THREE MINUTES

Range in rate for 9 minutes		*N*	Average change	CR
Fast				
Art	2800 up	42	+ 2 words	.07
History	2500 up	52	+28 words	1.09
Average				
Art	2200–2799	110	−39 words	3.03
History	1800–2499	178	−51 words	5.60
Slow				
Art	2199 down	157	−13 words	1.01
History	1799 down	103	−13 words	1.15

to have even more difficulty—but the results seem to indicate that poor readers plod ahead more-or-less inflexibly in their word-by-word reading without the difficulties of progressing comprehension much affecting their rate (the difference is not statistically significant).

TABLE VII.—AVERAGE INCREASE IN RATE FOR FICTION DURING THIRD THREE MINUTES OVER FIRST THREE MINUTES AT VARIOUS READING LEVELS

	9 minute rate change	*N*	Average increase, words	CR
Very fast	3200 up	43	+310	7.30
Fast	3000–3199	28	+163	6.08
Above average	2300–2999	132	+ 69	5.98
Below average	2299 down	170	+ 17	1.35

A comparison of fast and slow readers on the fiction selection also shows an interesting differentiation in rate pattern over successive intervals. The comprehension difficulty level of this selection is very easy; the typical reader averages ninety-four per cent of the questions

correct. Table VII shows that slow and even average readers tend to read fiction at a steady rate over successive intervals and not in the pattern of increasing rates as shown in Fig. 1.

As total rate for nine minutes increases beyond twenty-two hundred words, this pattern of increasing rate with progress through the story begins to show up. Further, as the total rate increases, this amount of gain during each period increases markedly. Fast readers of fiction seem able to benefit by comprehension of the story read thus far in foretelling or otherwise facilitating the process of comprehension in the later parts of the story. As compared to unit-by-unit comprehension in steady pace, this ability to benefit (increased rate) from present comprehension in reading later sections of the story must represent a higher level of reading ability.

CONCLUSIONS

These comparisons of differences in rate pattern at various levels of reading performance for a given selection would seem to indicate that they are not simple quantitative differences but rather differences in pattern or level of performance. There was no evidence to indicate clean-cut, absolute differentiation of these levels; rather there is often overlapping and a gradual shift to higher levels. These levels behave much as do higher level skills in typing, telegraphy, etc.

The lowest level of skill found is represented by word-by-word reading; a steady plowing of each line, little affected by comprehension difficulty. Different subject fields, varying greatly in difficulty, are read at comparable rates; and rate over successive time intervals remains remarkably constant.

A higher level of reading skill is demonstrated by most adult readers in their adjustment of rate to comprehension difficulty level; the fiction, art, history and geology selections varied greatly in difficulty and were read at different total rates. It is further characteristic of this group that as they found the assimilation in the non-fiction selections to be difficult, they slowed down in order to insure this assimilation. Flexibility in reading in adjusting to comprehension progress is the main characteristic of this level of ability.

A still higher level of reading skill seemed to be present in some of the faster readers of fiction—an ability to use the meaning of the story read thus far in facilitating the comprehension of later sections of the story. Such readers of fiction seem to be able to shift into an "overdrive" level once they get under way in a story.

The studies of rate change discussed here are for periods of nine minutes of reading. Extrapolation of these curves for any time periods beyond should be done with caution. The results of the present study indicate that the present practice of describing people's rates in terms of short rate tests is not descriptive of their rates over longer periods—such rates may be faster or slower and the direction of change will not be the same for all readers on the same selection. Also, the practice of describing an individual's rate in words-per-minute when this average is based on different time intervals is misleading. Average rate for an individual during two, four, ten or thirty minutes of reading from the same text will not be the same. It is quite disheartening to a student to hear that he reads three hundred words per minute on a three-minute history test, yet to find he cannot read twenty pages in his assignments in an hour.

III. USE OF READING AIDS IN ACADEMIC MATERIAL

A higher-level reading skill greatly emphasized in how-to-study work is the use of headings in non-fiction material. Since these headings indicate the ideas to be developed, their effective use should make for more efficient reading.

In an attempt to discover the effect of paragraph headings on the reading rate and comprehension accuracy of college students, one class section read the Russian history selection with headings included and the Canadian history selection with headings deleted. A second class section reversed this order and read Russian history "unheaded" and Canadian history "headed." These two history selections from the same editorial source have been shown above to be quite homogeneous, so the results of both tests have been combined in Table VIII.

TABLE VIII.—COMPARISON OF READING PERFORMANCE ON HEADED AND UNHEADED SELECTIONS

Test	Rate for nine minutes		Comprehension accuracy	
	Mean	SD	Mean	SD
Headed	2279	488	58.5	15.10
Unheaded	2266	419	57.7	16.45
$N = 62$	CR = .16		CR = .31	

It will be noted that there are but slight differences between the means and standard deviations for the headed and unheaded selections, for both rate and comprehension accuracy. The critical ratios are not statistically significant. The average college student apparently does not take advantage of one of the most useful devices in the study of textbook material.

The other studies reported in this paper have shown that certain higher-level reading skills do appear in some of the individuals studied, *i.e.*, flexibility in adjustment to comprehension difficulty and ability to facilitate comprehension from context. On the other hand, this last study indicates that few college students (representative of superior readers in the total population) have learned the higher-level technique of using headings as a means of reading textbook material most effectively. Obviously, such a lack of study skill indicates a definite instructional need of students.

FACTORS RELATED TO THE COLLEGE ACHIEVEMENT OF HIGH-APTITUDE STUDENTS WHO FAIL OF EXPECTATION AND LOW-APTITUDE STUDENTS WHO EXCEED EXPECTATION[1]

J. R. GERBERICH

University of Connecticut

INTRODUCTION AND METHOD OF INVESTIGATION

Two groups of students have been primarily responsible for the relatively low relationships found to exist between scores on prognostic college entrance tests and general scholastic success during the first term or semester of college attendance. These two groups consist of students whose test scores apparently indicate that they have superior qualifications for college work but who fall far below expectation in their scholastic performance and those students who, despite a seemingly low degree of aptitude for college work, far exceed expectation in their scholastic achievement. It is, of course, rather obvious that a degree of relationship approaching closely to that indicated by a correlation coefficient of +1.00 would be impossible of attainment with present measuring instruments and school organization, and also that no one not believing in a deterministic form of education could wish for such a situation. However, information concerning the factors associated with the discrepancies existing between prognostic test scores and college scholarship should be of value to college workers in the guidance and personnel fields.

The study reported here had as its aim the determination of how significant certain personality traits may be in effecting the disparity between college aptitude and achievement. No attempt was made to measure the personality traits themselves. Instead, the attempt was made to isolate certain reactions of the students to the college situation which certainly are influenced largely by the underlying personality make-up of the individual, and so to determine the operation of personality through a functional approach.

All freshman entrants to the University of Arkansas who took the American Council Psychological Examination at entrance and who completed at least the first semester of attendance during the years

[1] The writer wishes to express sincere appreciation to Mr. Fred L. Kerr, Registrar at the University of Arkansas, for permission to use the records in his office and for his active assistance in the collection of the follow-up data.

1929–1930, 1930–1931, and 1931–1932 composed the group basic to this investigation. Pearson product-moment coefficients of correlation between scores on the psychological test and measures of general scholastic success[1] for the first semester were .463, .575, and .546, respectively, for the years 1929–1930 to 1931–1932. These coefficients indicate that experience with the predictive power of this widely used entrance test at the University of Arkansas has been typical of experience elsewhere.[2]

Selection of the experimental students was made in terms of the degree to which first-semester scholastic achievement deviated from expectancy based on psychological test scores. Mean scores were computed for each row and column of the three scattergrams relating test scores to first-semester scholarship and smoothed once. Those students whose deviations from the smoothed means were more than one standard deviation in opposite directions—above the mean in intelligence and below the mean in scholarship, or below the mean in intelligence and above the mean in scholarship—were chosen as the experimental group. It is apparent that the students so selected were those whose scholastic success for the first semester of University attendance was in greatest disharmony with their scores on the prognostic examination.

Control groups were chosen to compare student for student with the experimental groups. For each student of superior aptitude who did poorly in scholarship, a control student of similar aptitude and like sex who achieved at or very close to the mean for his ability group was selected. Likewise, a control student of low aptitude who achieved near the mean for his ability level was paired with an experimental student of low aptitude who made a high scholastic record.

Throughout the remainder of this report, the letters "*E*" and "*C*" will be used to designate experimental and control groups, respectively. In like manner, the letters "*A*" and "*B*" will be used to designate, respectively, the students of high aptitude and those of low aptitude. The following, therefore, will be used hereafter to designate the four groups:

---

[1] A weighted grade point average (G.P.A.) was obtained by the use of the following values for semester marks: *A*6, *B*4, *C*2, *D*0, *E* −1, and *F* −2.

[2] Thurstone, L. L. and Thurstone, Thelma Gwinn: "The 1929 Psychological Examination." *Educational Record*, Vol. XI, April, 1930, pp. 101–128 and similar articles by the same authors concerning the 1930 and 1931 tests in *Educational Record*, Vol. XII, April, 1931, pp. 160–178 and Vol. XIII, April, 1932, pp. 121–136.

E-A Experimental students of high aptitude and low scholastic standing.
C-A Control students of high aptitude and average scholastic standing for their ability levels.
E-B Experimental students of low aptitude and high scholastic standing.
C-B Control students of low aptitude and average scholastic standing for their ability levels.

A questionnaire was sent to each experimental and control student involved in the investigation. The questionnaire consisted in the main of two types of items: (1) Those requiring affirmative or negative responses to questions concerning study habits, extra-curricular participation, reactions to the college situation, and the like,[1] and (2) Those requiring a statement of the average number of hours weekly devoted to each of a number of activities. The students were instructed to answer all items with reference to their first semester of attendance in the University.

Table I shows the number of experimental and control students involved in the study and the number of paired responses received. Unless responses were received from both members of a pair, the final tabulations included neither member of the pair. The high percentage of responses received was in part due to the fact that a few substitute control students were used when the control students first selected failed to respond. However, of the ninety-four control students, only ten were of this latter group.

TABLE I.—NUMBER OF PAIRED EXPERIMENTAL AND CONTROL STUDENTS AND NUMBER OF PAIRED RESPONSES RECEIVED

Group	1929		1930		1931		Totals	
	Number of pairs	Number of pairs responding	Number of pairs	Number of pairs responding	Number of pairs	Number of pairs responding	Number of pairs	Number of pairs responding
A	22	18	15	13	25	24	62	55
B	19	17	12	12	10	10	41	39

As shown by Table I, fifty-five of the sixty-two pairs of students having high aptitude completed the questionnaires and thirty-nine

[1] Acknowledgment is made to Herbert A. Toops, from whose Study Performance Test, Form *A*, Test 7, suggestions for some of these questions were obtained.

of the forty-one pairs of low-aptitude students did likewise. Thus, questionnaire returns were received from ninety-one per cent of the pairs, so that the findings are based on a very high percentage of the students falling within the scope of the investigation.

ENVIRONMENT, HEALTH, AND STUDY HABITS

Table II represents data concerning the differences in percentage of affirmative responses to questionnaire items given by the experimental and control groups of both high and low scholastic aptitude. Evidence concerning the significance of these differences is given for the three comparisons in terms of each difference divided by its probable error.[1] In the first two comparisons, differences are shown between each experimental group and its respective control group. The third comparison shows the differences between the two experimental groups. Of these last two groups, scholastic standings showed the *E-A* group to be the lower ranking despite the lower aptitude for college work shown by the *E-B* group.

The lack of quotients indicating the difference to be more than four times its probable error in the first two comparisons shows that the experimental and control groups did not differ by a statistically significant amount on any item. However, many of the quotients are sufficiently large as to make the chances of a real difference well over ninety in one hundred. Two significant differences appear between the affirmative responses given by the two experimental groups, however, while many others closely approach significance.

Interpretation of the results shown in Table II is dependent upon a knowledge of the items making up the questionnaire, for differences in methods of stating the items were such that a greater affirmative response to one item might be comparable to a smaller affirmative response to another. Table III lists the twenty-five items making up this section of the questionnaire and reproduces the columns of the preceding table which show the quotients of the differences and their probable errors. The last column of the table lists the scholarship group consistently giving the greater percentage of affirmative responses to each item.

Of the twenty-five items, complete consistency in the direction of differences between affirmative responses for the three comparisons exists for fifteen items, while for four other items two consistent

---

[1] Holzinger, Karl J.: *Statistical Methods for Students of Education.* Ginn, 1928, pp. 248–250.

TABLE II.—DIFFERENCES BETWEEN AFFIRMATIVE RESPONSES TO QUESTIONNAIRE ITEMS GIVEN BY PAIRED GROUPS IN THREE DIFFERENT COMPARISONS

Item number	Number *E-A*, *C-A*	*E-A*		*C-A*		Number *E-B*, *C-B*	*E-B*		*C-B*		*E-A* versus *C-A*			*E-B* versus *C-B*			*E-A* versus *E-B*		
		Proportion affirmative	PE_p	Proportion affirmative	PE_p		Proportion affirmative	PE_p	Proportion affirmative	PE_p	Difference[1]	PE_d	Diff. / PE_d	Difference[1]	PE_d	Diff. / PE_d	Difference[1]	PE_d	Diff. / PE_d
1	55	.255	.039	.218	.038	39	.282	.049	.282	.049	−.037	.055	0.67	.000	.069	0.00	.027	.063	0.43
2	55	.218	.038	.091	.026	39	.154	.038	.103	.033	−.127	.046	2.76	.051	.051	1.00[2]	−.064	.054	1.19
3	55	.182	.035	.091	.026	39	.103	.033	.051	.024	−.091	.044	2.07	.052	.040	1.30[2]	−.079	.048	1.65
4	55	.891	.028	.873	.030	38	.947	.024	.921	.030	−.018	.042	0.43	.026	.038	0.68[2]	.056	.040	1.40[2]
5	55	.964	.016	.964	.016	39	.923	.028	.949	.024	.000	.024	0.00	−.026	.037	0.70	−.041	.044	0.93
6	55	.418	.045	.291	.042	38	.158	.040	.316	.051	−.127	.061	2.08	−.158	.065	2.43	−.260	.052	5.00
7	53	.509	.047	.453	.047	38	.368	.053	.553	.055	−.056	.065	0.86	−.185	.077	2.40	−.141	.059	2.39
8	51	.373	.046	.275	.042	37	.351	.053	.378	.054	−.098	.062	1.58	−.027	.076	0.36	−.022	.060	0.37
9	54	.167	.034	.111	.028	39	.128	.036	.308	.050	−.056	.045	1.24	−.180	.061	2.95	−.039	.049	0.80
10	54	.815	.036	.907	.027	38	.974	.018	.947	.024	.092	.045	2.04[2]	.027	.030	0.90[2]	.159	.034	4.68[2]
11	54	.482	.046	.482	.046	38	.632	.053	.579	.054	.000	.065	0.00	.053	.076	0.70[2]	.150	.059	2.54[2]
12	55	.309	.042	.218	.038	38	.447	.055	.395	.053	−.091	.057	1.60	.052	.076	0.68[2]	.138	.061	2.26[2]
13	55	.655	.043	.727	.040	39	.846	.038	.692	.050	.072	.059	1.22[2]	.154	.063	2.44[2]	.191	.051	3.75[2]
14	54	.296	.042	.315	.042	39	.359	.052	.308	.050	.019	.060	0.32[2]	.051	.072	0.71[2]	.063	.059	1.07[2]
15	55	.527	.045	.473	.045	39	.333	.051	.410	.053	−.054	.064	0.84	−.077	.074	1.04	−.194	.059	3.29
16	55	.273	.040	.127	.030	39	.154	.038	.154	.038	−.146	.051	2.86	.000	.055	0.00	−.119	.056	2.11
17	55	.218	.038	.218	.038	39	.308	.050	.359	.052	.000	.053	0.00	−.051	.072	0.71	.090	.063	1.43[2]
18	55	.091	.026	.146	.032	39	.180	.042	.180	.042	.055	.042	1.31[2]	.000	.059	0.00	.089	.049	1.82[2]
19	54	.074	.024	.055	.020	39	.026	.016	.128	.036	−.019	.032	0.59	−.102	.040	2.55	−.048	.030	1.60
20	54	.037	.018	.109	.028	39	.103	.033	.026	.016	.072	.034	2.12[2]	.077	.037	2.08[2]	.066	.038	1.74[2]
21	54	.296	.042	.291	.042	39	.180	.042	.282	.049	−.005	.059	0.08	−.102	.064	1.59	−.116	.068	1.71
22	52	.654	.045	.636	.044	39	.590	.053	.692	.050	−.018	.063	0.29	−.102	.073	1.40	−.064	.075	0.85
23	54	.741	.040	.907	.027	39	.923	.028	.872	.036	.166	.049	3.39[2]	.051	.047	1.09[2]	.182	.059	3.08[2]
24	53	.774	.038	.815	.036	39	.949	.024	.846	.038	.041	.053	0.77[2]	.103	.045	2.29[2]	.175	.063	2.78[2]
25	53	.642	.045	.849	.033	38	.718	.049	.564	.053	.207	.063	3.29[2]	.154	.072	2.14[2]	.076	.072	1.06[2]

[1] Positive differences show a greater proportion of affirmative response for the group attaining the higher scholastic average—*C-A*, *E-B*, and *E-B*, respectively, for the first, second, and third comparison; negative differences show a greater proportion of affirmative response for the group making lower scholastic averages—*E-A*, *C-B*, and *E-A*.

[2] Differences, in terms of their probable errors, which show a greater affirmative response by the more successful scholastic group.

differences are accompanied by one comparison showing no difference whatsoever. Consistency exists for at least sixty per cent of the twenty-five items, and, if the four comparisons mentioned above are considered consistent, this percentage is raised to seventy-six. In either event, the percentage is much greater than would result from pure chance. Therefore, although few of the individual comparisons are statistically significant, the consistency with which they are found to agree in direction for similar comparisons may well be taken as further evidence concerning the degree of meaning which may be attached to them.

The most meaningful consistent differences found show the groups of lower scholastic achievement to have experienced greater difficulty in paying attention in class, to have had greater difficulty in taking notes and reading graphs, to have pledged fraternities or sororities in greater numbers, to have participated more in football and intramural athletics, and to have found university work more difficult than high-school work to a greater degree than was true of the higher-ranking scholarship groups.

On the other hand, the students of the higher scholarship groups usually handed themes and papers in promptly, crammed for examinations, had a good place to study, used the library for study, participated in freshman basketball, liked most of their courses and instructors, and felt that they got a "square deal" in all of their classes in a greater percentage of the cases than was true of the less successful students paired with them.

EXPENDITURE OF TIME

Tables IV and V present data comparable to those above for the other type of questionnaire responses—average number of hours per week spent in each of several major types of activity. One of the questionnaire items in this section was labeled "Miscellaneous," in order to allow the student a place in which to list those hours of the one hundred sixty-eight per week which he could not properly classify under one of the listed types of activity. The results from this miscellaneous classification have not been tabulated. For three of the classifications used—work, athletics, and fraternity or sorority—average time expenditures per week proved an unreliable measure because of the fact that many students spent no time in these three activities. Therefore, a less satisfactory measure was used—the proportion of each group participating one hour or more per week in each of these three types of activity.

TABLE III.—ITEMS OF QUESTIONNAIRE AND CONSISTENCY OF THREE COMPARISONS BETWEEN GROUP OF RELATIVELY HIGH AND LOW SCHOLASTIC ACHIEVEMENT

Item number	Question	*E-A* versus *C-A*	*E-B* versus *C-B*	*E-A* versus *E-B*	Scholastic group showing consistently greater affirmative response
1	Did you usually feel tired when you got up in the morning?[3]	0.67	0.00	0.43[1]	
2	Did you have frequent sick spells or periods of feeling bad?	2.76	1.00[1]	1.19	
3	Did you miss many classes because of illness?	2.07	1.30[1]	1.65	
4	Did you have all the textbooks you needed for your courses?	0.43	0.68[1]	1.40[1]	
5	Did you have all the necessary equipment (notebooks, pens, pencils, etc.), for your classes?	0.00	0.70	0.93	Lower[2]
6	Did you have difficulty in paying attention in class?	2.08	2.43	5.00	Lower
7	Did you have difficulty in taking class notes which would serve as a good foundation for later reviews?	0.86	2.40	2.39	Lower
8	Did you have difficulty in outlining or taking notes on reading assignments?	1.58	0.36	0.37	Lower
9	Did you have difficulty in interpreting the tables or graphs you encountered in your lessons?	1.24	2.95	0.80	Lower
10	Did you usually hand in themes and term papers on time?	2.04[1]	0.90[1]	4.68[1]	Higher
11	Did you cram for examinations?	0.00	0.70[1]	2.54[1]	Higher[2]
12	Did you get nervous and excited when taking final examinations?	1.60	0.68[1]	2.26[1]	
13	Did you have a good, quiet place to study where you lived?	1.22[1]	2.44[1]	3.75[1]	Higher
14	Did you do much of your studying in the library?	0.32[1]	0.71[1]	1.07[1]	Higher
15	Did you pledge a fraternity or a sorority during your first semester?	0.84	1.04	3.29	Lower
16	Did you live in a fraternity or sorority house at least three months of the first semester?[3]	2.86	0.00	2.11	Lower[2]
17	Did you do any outside work for pay or for board and room?	0.00	0.71	1.43[1]	
18	Did you have an automobile which you could use regularly?	1.31[1]	0.00	1.82[1]	Higher[2]
19	Did you play freshman football?	0.59	2.55	1.60	Lower
20	Did you play freshman basketball?	2.12[1]	2.08[1]	1.74[1]	Higher
21	Did you participate in intramural sports?	0.08	1.59	1.71	Lower
22	Did you find university work much harder than high school work?	0.29	1.40	0.85	Lower
23	Did you like most of the subjects you took the first semester?	3.39[1]	1.09[1]	3.08[1]	Higher
24	Did you like most of your instructors the first semester?	0.77[1]	2.29[1]	2.78[1]	Higher
25	Did you feel that you got a "square deal" in all of your classes?	3.29[1]	2.14[1]	1.06[1]	Higher

[1] Greater proportion of affirmative response for group attaining higher scholarship.

[2] One zero and two consistent differences.

[3] Statement revised to conform with method of tabulating response to the item which, as originally used, required a check mark in front of one response among several listed.

TABLE IV.—DIFFERENCES BETWEEN TIME DISTRIBUTIONS OF PAIRED GROUPS IN THREE DIFFERENT COMPARISONS

Activity	*E-A* (*N* = 54)		*C-A* (*N* = 54)		*E-B* (*N* = 37)		*C-B* (*N* = 37)		*E-A* versus *C-A*			*E-B* versus *C-B*			*E-A* versus *E-B*		
	Mean	PE_m	Mean	PE_m	Mean	PE_m	Mean	PE_m	Difference	PE_d	$\frac{Diff.}{PE_d}$	Difference[1]	PE_d	$\frac{Diff.}{PE_d}$	Difference[1]	PE_d	$\frac{Diff.}{PE_d}$
Sleep	53.61	0.68	54.54	0.53	55.58	0.89	53.20	0.66	0.93	0.87	1.07[2]	2.38	1.10	2.16[2]	1.97	1.12	1.76[2]
Meals	9.91	0.38	10.59	0.45	10.57	0.55	10.08	0.48	0.68	0.59	1.15[2]	0.49	0.73	0.67[2]	0.66	0.67	0.99[2]
Classroom and laboratory	21.17	0.55	21.72	0.47	23.96	0.79	23.26	0.92	0.55	0.72	0.76[2]	0.70	1.21	0.58[2]	2.79	0.96	2.91[2]
Studying	19.13	0.88	23.94	0.87	27.12	1.38	25.93	1.24	4.81	1.24	3.88[2]	1.19	1.90	0.63[2]	7.99	1.64	4.87[2]
Reading	7.78	0.48	7.54	0.47	5.08	0.31	5.30	0.56	−0.24	0.67	0.36	−0.22	0.64	0.34	−2.70	0.57	4.74
Movies, etc.	10.98	0.54	8.39	0.64	8.45	0.83	8.76	0.75	−2.59	0.99	2.62	−0.31	1.12	0.28	−2.53	0.99	2.56
	Proportionate part	PE_p	Proportionate part	PE_p	Proportionate part	PE_p	Proportionate part	PE_p	Difference	PE_d	$\frac{Diff.}{PE_d}$	Difference	PE_d	$\frac{Diff.}{PE_d}$	Difference	PE_d	$\frac{Diff.}{PE_d}$
Work	.259	.040	.259	.040	.351	.053	.405	.055	.000	.057	0.00	−.054	.076	0.71	.092	.066	1.39[2]
Athletics	.722	.041	.759	.039	.649	.053	.703	.051	.037	.057	0.65[2]	−.054	.074	0.73	−.073	.067	1.09
Fraternities and sororities	.537	.046	.463	.046	.405	.055	.405	.055	−.074	.065	1.14	.000	.077	0.00	−.132	.072	1.83

[1] Positive differences show a greater time expenditure for the group attaining the higher scholastic average—*C-A*, *E-B*, and *E-B*, respectively, for the first, second, and third comparisons; negative differences show a greater time expenditure for the group making lower scholastic averages—*E-A*, *C-B*, and *E-A*.

[2] Differences, in terms of their probable errors, which show greater time expenditures by the more successful scholastic group.

The differences between the time expenditures of the two experimental and their respective control groups and also between the two experimental groups are shown in Table IV in relation to the probable errors of the differences.[1] The quotients of the differences and their probable errors approach closely to significance in only one of the eighteen comparisons between experimental and control groups—amount of time spent in study by the high-aptitude experimental and control groups.

The last columns of the table show findings for the two groups having the greatest variability—students of high aptitude who do low grade scholastic work and students of low aptitude who do high grade scholastic work. Among the nine comparisons, two show significant differences. The superior scholastic success of the low-aptitude students is apparently accompanied by a significantly greater amount of time spent in study and a significantly smaller amount of time spent in optional types of reading than is the low scholarship of the group having higher aptitude.

Table V summarizes the findings of the preceding table and indicates the degree of consistency found in the various methods of time expenditure for the three comparisons. Of the nine types of activity, six show complete consistency in the direction of the differences found

TABLE V.—CONSISTENCY OF DIFFERENCES IN TIME EXPENDITURE BETWEEN GROUPS OF RELATIVELY HIGH AND LOW SCHOLASTIC ACHIEVEMENT

Activity	*E-A* versus *C-A*	*E-B* versus *C-B*	*E-A* versus *E-B*	Scholastic group showing consistently more time spent (or greater participation)
Sleep	1.07[1]	2.16[1]	1.76[1]	Higher
Meals	1.15[1]	0.67[1]	0.99[1]	Higher
Classroom	0.76[1]	0.58[1]	2.91[1]	Higher
Studying	3.88[1]	0.63[1]	4.87[1]	Higher
Reading	0.36	0.34	4.74	Lower
Movies, etc	2.62	0.28	2.56	Lower
Work[2]	0.00	0.71	1.39[1]	
Athletics[2]	0.65[1]	0.73	1.09	
Fraternities and sororities[2]	1.14	0.00	1.83	Lower (?)

[1] Greater amount of time spent (or proportion of participation) by the group attaining higher scholarship—*C-A*, *E-B*, and *E-B*, respectively, for the first, second, and third comparisons.

[2] Based on proportion of group participating, rather than on the amount of time spent.

[1] *Ibid.*, pp. 235–237.

from the three comparisons for each. One other activity shows consistency for two comparisons and a zero difference for the third. More important, perhaps, is the fact that all six of the comparisons based on actual time distributions are wholly consistent. It is probable, then, because of the consistency shown by these findings, that the nature of the obtained differences should be given more consideration than the amount of the difference for any single comparison might warrant.

The above evidence supports the conclusion that the more successful students spend more time than the relatively unsuccessful students in sleep, at meals, in the classroom or laboratory, and in studying. The less successful students appear to spend more time in voluntary or optional reading not connected with class work, at the movies, and in other forms of entertainment, and perhaps are more frequently pledged to campus fraternal groups.

SUBSEQUENT SCHOLASTIC SUCCESS

Inasmuch as students have so far in this paper been judged as successful or unsuccessful in terms of their first semester scholarship in college, it is important that evidence be introduced to show whether first semester scholarship is a reliable indication of their success during later years in college. Therefore, the data of Table VI are presented to indicate the subsequent college success of both the experimental and the control student in terms of the two most widely used criteria of success—scholastic standing and persistence in attendance. To obtain the data here reported, the collegiate careers of the students were followed through to the close of the school year 1935–1936.

It is immediately apparent that most of the differences shown in the last columns of this table are much more than four times their probable errors. The quotients for cumulative grade-point averages range from 9.30 for the *E-B* vs. the *C-B* groups to 14.44 for the *E-A* vs. the *C-A* groups. Therefore, the differences in first-semester scholarship are maintained as real and significant differences when the students' entire scholastic records are considered.

The quotients show that for hours of credit earned and semesters of college attendance at the University of Arkansas, the *E-A* group differed significantly from both the *C-A* and the *E-B* groups, but that the differences between the *E-B* and *C-B* groups are negligibly small. It appears, then, that differences in first-semester scholarship tend to carry on in somewhat diminished degree in comparisons between students of high aptitude but low initial performance and students of high aptitude and average performance for their ability levels and also

TABLE VI.—DIFFERENCES IN LATER SCHOLASTIC SUCCESS AND PERSISTENCE OF PAIRED GROUPS IN THREE DIFFERENT COMPARISONS

Criterion	*E-A* (*N* = 55)		*C-A* (*N* = 55)		*E-B* (*N* = 39)		*C-B* (*N* = 39)		*E-A* versus *C-A*			*E-B* versus *C-B*			*E-A* versus *E-B*		
	Mean	PE_m	Mean	PE_m	Mean	PE_m	Mean	PE_m	Difference[1]	PE_d	Diff./PE_d	Difference[1]	PE_d	Diff./PE_d	Difference[1]	PE_d	Diff./PE_d
First semester GPA....	0.339	0.086	2.802	0.060	3.407	0.094	1.112	0.062	2.463	0.105	23.46	2.295	0.113	20.31	3.068	0.128	23.97
Cumulative GPA.....	0.911	0.086	2.716	0.090	2.792	0.123	1.382	0.089	1.805	0.125	14.44	1.410	0.152	9.30	1.881	0.150	12.54
Hours credit.....	50.09	3.24	88.00	3.42	79.62	4.36	73.98	4.80	37.91	4.71	8.05	5.64	6.48	0.87	29.53	5.43	5.44
Semesters attended.....	3.73	0.20	5.62	0.21	5.05	0.26	4.90	0.28	1.89	0.29	6.52	0.15	0.38	0.39	1.32	0.32	4.13

[1] All differences show higher scholarship or higher standing on the persistence criteria for the group attaining the higher initial scholastic average—*C-A*, *E-B*, and *E-B*, respectively, for the first, second, and third comparisons.

in comparisons between students of high aptitude and low initial achievement and those of low aptitude but high achievement. However, there is evidence to indicate not only that the low-aptitude students whose initial achievement is average for their ability levels persist in school nearly as long as their classmates of similar ability and higher initial achievement, but also that their persistence is actually greater than that of the group of students vastly superior in aptitude but low relatively in initial scholastic success.

SUMMARY AND CONCLUSIONS

In the study which is basic for this report, four groups of students entering the University of Arkansas as freshmen in 1929, 1930, and 1931—students who vastly surpassed and who greatly fell below the scholastic expectation warranted by their low and high scores, respectively, on the American Council Psychological Examination and control groups of high and of low aptitude students whose first-semester scholarship was close to expectation—filled out questionnaires with information concerning their first-semester careers in the institution. All four groups of students were also followed through their careers at the University of Arkansas with respect to scholarship and persistence.

Three types of comparisons between group reactions were made: (1) Between high-aptitude, low-scholarship and high-aptitude, high-scholarship students, (2) between low-aptitude, high-scholarship and low-aptitude, low-scholarship students, and (3) between high-aptitude, low-scholarship and low-aptitude, high-scholarship students. Questionnaire responses of two types were compared for these groups: (1) Affirmative responses to twenty-five questions dealing with study habits, extra-curricular participation, and related issues, and (2) hours of time per week spent in various types of activities.

Of the seventy-five group comparisons based on the twenty-five questions, few of the differences were statistically significant. However, consistency of differences for the three possible comparisons on each question existed for sixty per cent of the items, a finding which seems to justify a belief that the findings are more reliable than the separate group comparisons would indicate. The low-scholarship groups were consistently found to have greater difficulty in paying attention in class, to have greater difficulty in using study techniques, to be fraternity or sorority pledges more frequently, to participate more often in freshman football and intramural athletics, and to recognize a greater degree of difference in difficulty between high-school and university work than was true for the high-scholarship

groups. The high-scholarship groups, on the other hand, more frequently than the low-scholarship groups, handed themes in promptly, crammed for examinations, had a good place to study, used the library for study, participated in freshman basketball, liked most of their classes and instructors, and felt that they got a "square deal" in their classes.

Data on time expenditures show that the high-scholarship students consistently spent more time in sleep, at meals, in the classroom, and studying, and less time at the movies and in optional reading than did the low-scholarship groups.

Although the relationships stated above are not shown by the data to be causally related necessarily, the evidence lends support to the belief that at least two of the factors contributing to the low scholarship of high-aptitude students are study techniques and time expenditure not most conducive to the attainment of high course marks, and that at least a partial explanation of the high marks made by low-aptitude students lies in careful study methods, attention to the demands made by instructors, and elimination of extra-class activities from their schedules.

IMPLICATIONS

The implications of the findings are not entirely clear. There are those who, believing that scholastic achievement of a high order is the major if not the sole aim of college attendance, might find solace in the situation shown to exist. There are others, however, who, believing that only a part of the student's education is obtained in the classroom and that highly desirable development in the student may result from non-academic activities (and consequently not receive consideration in his scholastic grade point average), may be disturbed by these findings, typical though they doubtless are. There are also some who, probably falling largely within the second group, might see in these findings an indication that the most frequently accepted criterion of college success—scholastic standing—measures only a portion of the outcomes accruing to attendance at a college or university. These persons might conclude, with some justification, that the sacrifice often made for high marks is too great a price for some students to pay, and that the sacrifice of highly able students through withdrawal or failure to return (in extreme cases by the "dropped for poor scholarship" route), because the institution fails to meet their needs adequately or to furnish them proper guidance, is too great a price for such students, the institution, and society to pay.

THE USE OF TESTS OF PERSISTENCE IN THE PREDICTION OF SCHOLASTIC ACHIEVEMENT

G. R. THORNTON

Purdue University

It has long been assumed that persistence is one of the factors determining achievement in school.* Recently experimental evidence has been presented to show that prediction of grades in college is improved by use of scores on objective tests of persistence. Howells[3] in 1933 reported that the scores on a battery of persistence tests yielded a coefficient of correlation of .44 with average grades for the preceding year for one hundred college students. The correlation between intelligence scores and grades for these same students was .51; and "the multiple correlation of a combination of intelligence and persistence scores with grades was .64." In 1938 Ryans[6] reported a correlation of .48 between a group test of persistence and honor-point ratios (that is, grades) for forty subjects. As pointed out elsewhere[8] there is some question concerning the validity of Ryans' results, since in the selection of the component parts of his test of persistence honor-point ratios were used as one of the criteria, and the forty subjects for whom the final correlation was reported are the same persons that made up the criterion group. Nonetheless, these reports lead one to expect to find appreciable correlation between achievement in school and scores on tests of persistence. Indeed, the above results have led the writer of a widely used textbook to assert that "a battery of tests may measure persistence rather well," and "The persistence score shows almost zero correlation with intelligence, but an appreciable positive correlation with academic achievement."[10]

If the above results are accepted at their face value, it would seem that persistence is one of the factors determining success in college and that tests of persistence might well be included as a part of the battery of tests used by various colleges and universities as entrance examinations. The results to be reported in the present paper, however, are less favorable to tests of persistence; and certain implications arising from a comparison of the present results with those of Howells and Ryans suggest the need for caution in the interpretation of the relationships reported between academic achievement and scores on tests of persistence.

* My interest in this problem was first aroused by Dr. W. R. Baller through the generous loan of an unfinished manuscript on the topic.

Before one can establish a relationship between persistence and academic achievement, it is necessary to have tests that measure persistence. Considerable evidence has been presented which indicates that the various tests which claim to measure persistence do not all measure the same thing. In fact, some of the more widely accepted tests show little or no correlation with each other.[5,7,2] Even in Ryans' group test of persistence, referred to above, the three component subtests show practically no correlation with each other. The coefficients reported by Ryans are .21, −.04, and .29. (A statistically significant coefficient for forty subjects would be .31 or above.) The highest of these coefficients, .29, is for the correlation between the subtests "endurance" and "study time" which, when used in essentially the same form as individual tests, yielded a correlation of only .09. These facts raise the question whether scores on the so-called tests of persistence should be considered measures of persistence.

In a factor analysis of a battery of tests which have been considered to be tests of persistence, the present author[7] found two factors which he labelled "*WD*" and "*Pl*," which seemed to resemble what has been called persistence; but these two factors were practically unrelated to each other. The original analysis was of correlations for scores of one hundred eighty-nine men and women students on twenty-two items. In a second analysis of part of the same data the intercorrelations of fifteen of the most significant items for the one hundred men who were most homogeneous in age were factorized. The *WD* and *Pl* factors appeared again. The results of this second analysis form the basis of the present report. In the *WD* factor the two items that were by far the most heavily weighted were the "shock" and "pressure" tests. These two tests were patterned closely after Howells' tests, and were equivalent to eighty per cent of his battery of persistence tests (referred to above). In the *Pl* factor, also, two tests were by far the most heavily weighted: a word building test and a "perceptual ability" test. In both of these tests the subject was set a verbal task that became increasingly difficult; he was permitted to continue working at the task as long as he wished, and his score was the time he continued to work.*

Since the *WD* factor appears to be what Howells considered to be persistence, and since the tests in the *WD* and *Pl* factors resemble those included in Ryans' group test of persistence, one should expect scores

* For description of the shock, pressure, word-building, and perceptual ability tests see [7], 5f and 9f.

on these factors to show appreciable correlation with academic achievement in college. The coefficients of correlations between these factors and academic achievement were calculated. Two measures of academic achievement were used: (1) The weighted average of all grades earned (to date) at the University of Nebraska (2) the grade earned in elementary psychology in which all the subjects were enrolled. The scores used to represent *WD* were the averages of *T* scores on the shock and pressure tests. For the *Pl* factor the scores used were the averages of *T* scores on the word building and perceptual ability tests. Coefficients were also calculated between the measures of academic achievement and the total of the *T* scores for *WD* and *Pl* (hereafter labelled simply *WD & Pl*). All correlations were calculated on the basis of results for ninety-eight subjects, because grades were not available for two of the one hundred men in the group. The correlations obtained are presented in Table I.

TABLE I.—CORRELATIONS FOR NINETY-EIGHT SUBJECTS

	Average grades	Psychology grade
Pl	.25	.18
WD	.17	.18
WD & Pl	.29	.25

Since the subjects were enrolled in different colleges within the University, and since the grade averages varied in the different colleges, it was thought that the correlation between grades and *WD & Pl* might be higher if only students within one college were included. The correlation between *WD & Pl* and grades for the forty-two students in the College of Arts and Sciences, however, was found to be only .31.

According to Fisher's test of significance, for ninety-eight subjects a coefficient of .20 is significant and one of .26 is very significant. While, therefore, some of the coefficients listed in Table I are statistically significant, the more striking fact is the smallness of the coefficients. Particularly striking is the small correlation between grades and *WD* (.17). Howells, it will be recalled, found a comparable coefficient of .44 for one hundred cases.

Various hypotheses might be suggested to account for the difference in the size of the coefficients reported here and those reported by Howells and Ryans. As indicated above, there is reason to believe that the correlation reported by Ryans is spuriously high; how much

too high can only be conjectured. It might be suggested that the tests in the present study were less reliable; but, although the reliabilities reported in the three studies are not directly comparable, what evidence there is seems to indicate that the tests in the present study were more rather than less reliable. (Ryans reports a test-retest reliability of .82 for his battery; Howells reports split-half reliabilities of .89 and .92 for his battery; in the study of one hundred eighty-nine subjects the present writer found split-half reliabilities of .945 and .956 for the shock and pressure tests, respectively.) Perhaps there were differences in the reliabilities of the grades or in what grades indicate in the different schools; but little evidence is available by which one can evaluate these possibilities. Again, it might be suggested that the differences are due to differences in the range and type of performances sampled in the three studies. This hypothesis, however, loses much of its force when one considers that the greatest discrepancy is between (1) the correlation reported for *WD* and average grades in the present study and (2) the correlation reported for Howells' battery and average grades. *WD* in the present study was based on scores on six pressure tests and two shock tests, and these same tests comprised eight of the ten units in Howells' battery.

The hypothesis that seems to be most fruitful is that the amount of correlation between grades and scores on persistence tests is a function of the degree to which the test situations are similar to school situations. Two aspects of the similarity are probably of greatest importance: (1) Similarity of the tasks performed in the test situations to those performed in school and (2) similarity of the social relationship of experimenter and subject to the relationship of instructor and student. In two of the three studies under consideration it is expressly stated that the subjects were students from classes in which the experimenter was the instructor; in the third study (Howells') the subjects were mostly students in elementary psychology classes, and were probably from the experimenter's own classes. Two considerations, however, would seem to indicate a lesser degree of influence of the instructor-student relationship in the present study: (1) The experimenter appears young and was a graduate student at the time of the experiment. (2) During the administration of the tests considerable attention was paid to maintaining standard test conditions—instructions were memorized or read, and rest periods were for the most part timed; this probably tended to make the experimental situation more impersonal and so to reduce the importance of the immediate social

factors in the situation. With the exception of the part of Ryans' battery called "study time," there appear to have been no important differences between the three studies in the degree of similarity of experimental tasks to school tasks. It is interesting to note, however, that in the present study the *Pl* factor, which was measured by scores on verbal tasks, shows somewhat higher correlation with grades than does the *WD* factor, which was measured by scores on non-verbal and unschool-like tasks.

The evidence at hand obviously is not adequate to establish the above hypothesis; and, indeed, the amount and reliability of the differences in degree of correlation for the explanation of which the hypothesis is suggested are scarcely sufficient to warrant extended discussion. The hypothesis has been suggested, and evidence for it presented, primarily because of an implication of possible significance; namely, that the best way to increase the value of objective personality tests as a means of predicting scholastic achievement is not by making the tests better measures of persistence, but rather by planning the test situations to resemble more closely the tasks and social relationships found in the classroom. The significance of this implication seems sufficient to warrant further studies to test the truth of the proposed hypothesis.

The hypothesis can be tested by comparing coefficients of correlations between grades or other measures of academic achievement and scores on objective tests when (1) the tasks in the test situations are varied from those closely resembling school tasks to those bearing little resemblance to school tasks, and (2) the relationship of experimenter to subject is varied in its degree of resemblance to the instructor-student relationship. The latter variation can be accomplished readily by having the experimenter in one series of tests be a fellow-student and in a second series an instructor. If the hypothesis is correct, the correlation between grades and test scores should increase (1) as the tasks in the tests become more like school tasks and (2) as the relationship between experimenter and subject becomes more like that between instructor and student.*

---

* A corollary of this is that test scores should be relatively better predictive devices for the prediction of grades that are determined on less objective bases, that is, when the social interaction between instructor and student plays a more important part in the determination of grades. The absolute size of the correlation in this latter case, however, would probably not be greater, since the increased degree of relationship might be offset by decrease in the reliability of the grades.

A further extension of the above hypothesis leads to the conclusion that the best test situation for determining a student's reactions to classroom tasks and classroom social relationships would be the classroom itself. This conclusion is supported by the fact cited by Williamson [9], p. 87 that "the average of all high-school grades continues to be the best single predictor of average grades in a number of colleges."* Grades, of course, reflect in large prat the scholastic ability of the student, for which we already have fairly satisfactory measuring devices; but grades undoubtedly reflect also other factors, such as interests, persistence and application, adequacy of emotional adjustment, adaptation to the classroom, docility and other attitudes toward instructors. These other factors are the ones for which we need measuring devices. Is it not possible that the best measure we can obtain of these other factors is the average of grades with an appropriate correction for ability?

Let us lump all these factors other than scholastic ability (as measured by present tests) under the heading "scholastic efficiency." An index of scholastic efficiency might be A/B, in which A represents average T score (or Hull score) for grades in a given school and B represents T score (or Hull score) for scholastic ability in comparison with other students in the same given school.† Such an index would have no value for predicting future grades in the same school, since one might just as well, and more easily, use the average grades for that purpose. It would have considerable value (as Peters has suggested) as an indication of the degree to which a student is achieving what he is capable of achieving. The value of such an index, so far as prediction is concerned, lies in its potential usefulness in predicting future performance at the time when the student is entering a new school—when he enrolls in a college or university, for example. The formula for predicting achievement in the new school might be: $X = AY/B$, in which X = future scholastic achievement, $A = T$ score for grades in the previous school, $B = T$ score for scholastic ability in terms of classmates in the previous school, and $Y = T$ score for scholastic ability in terms of classmates in the new school.

The above formula overlooks what is probably an important psychological factor determining scholastic achievement; namely, that a student tends to think of himself as an average student or as a

* See also [1], p. 125.

† This formula has been used by C. C. Peters[4] as an "accomplishment quotient," and similar formulas have been used by others.

superior student, etc.; and this subjective standard tends to determine the degree of effort exerted. That is, when a student, whose level of aspiration is to be an average student, finds upon entering a new school that it is more difficult to earn average grades than it was previously, he probably will put forth more effort in order to attain his standard of achievement. The importance of the student's level of aspiration in determining his achievement and the degree to which his level of aspiration will be changed by the changed requirements of the new school situation, will probably vary widely from individual to individual. As a preliminary guess concerning the influence of the level of aspiration for the average student, we might assume that its importance as a determiner is approximately one-half that of the other factors determining the scholastic efficiency of the student. Assuming that the average of grades earned in the previous school is a rough measure of level of aspiration, our formula becomes: $X = \frac{2AY}{3B} + \frac{A}{3}$. (The symbols have the same meaning as above.)

The usefulness of this formula for the prediction of grades must be determined by empirical test. It is to be expected that its validity will be less for certain extreme cases. Capable students who have been bored by the work in a school where the requirements are low and who, consequently, have shown a low degree of scholastic efficiency, may upon entering stiffer competition in a new school change radically both in their level of aspiration and in their scholastic efficiency. On the other hand, capable students whose index of scholastic efficiency is about 1.00, because they are at the top of their classes in both ability and achievement,* upon entering stiffer competition also may earn a higher index of scholastic efficiency. The modified formula above, which takes into account the level of aspiration, would compensate in part for the error in the case of the latter students, but not in the case of the former students. Likewise, students who have unusually strong need to attain a given level of achievement would not be expected to conform to the formula, although with greater information concerning the relationships involved it may be possible in the future to modify the formula to take into account differences in the strength and rigidity of the level of aspiration. Students, who because of

* Because of the method of calculating the index of scholastic efficiency, the most able student cannot earn an index of more than 1.00, whereas the less able student can earn an index of well over 1.00. Peters [4], p. 104ff suggests a correction formula to compensate for this deficiency.

intense need have been achieving far in excess of their level of ability, upon entering stiffer competition where it is no longer possible for them to attain their standard of performance, may "give up" and so show a sudden drop in scholastic efficiency.

The immediate utility of this prediction formula is sharply limited by the fact that few secondary schools have and make available the data needed for calculating the index of scholastic efficiency. This lack, it is to be hoped, will be remedied in the near future, since the guidance value of such an index, even in the secondary school itself, would more than justify the cost of gathering the data.

In the absence of data from which to calculate an index of scholastic efficiency, it would seem worth while to experiment with various objective personality tests in the hope of developing measures of at least some of the factors other than ability which determine scholastic achievement. That it is possible to develop such measures is suggested by the results of Howells and Ryans and, to a lesser extent, by the results reported in this paper. It is this writer's belief, however, that tests are more likely to prove useful for prediction of scholastic achievement if they are designed to sample performances involving more general aspects of student reactions rather than persistence, suggestibility, coöperativeness, etc., separately.

SUMMARY

Coefficients of correlation between grades and scores on tests designed to measure persistence are reported and compared with results from similar investigations by Howells and Ryans. An hypothesis is presented to account for discrepancies in the results of the three investigations, and the implications of this hypothesis for the problem of predicting grades are discussed somewhat in detail. These implications lead to the hypothesis that objective personality tests will have value as devices for predicting scholastic achievement in proportion to the degree to which the tests approach in similarity the tasks and social relationships of the actual classroom. In harmony with this hypothesis a formula is suggested for the prediction of grades in a new school on the basis of scholastic ability plus an "index of scholastic efficiency" and the level of aspiration displayed in a previous school.

REFERENCES

1. Harris, Daniel: "Factors affecting college grades: A review of literature, 1930–1937." *Psychol. Bull.*, Vol. XXXVII, March, 1940, pp. 125–166.

2. Hartshorne, M., May, M. A., and Maller, J. B.: *Studies in service and self-control.* New York: Macmillan, 1929.
3. Howells, T. H.: "An experimental study of persistence." *J. abnorm. soc. Psychol.*, Vol. XXVIII, 1933, pp. 14–29.
4. Peters, C. C.: A method for computing accomplishment quotients on the high-school and college levels." *J. educ. Research*, Vol. XIV, 1926, pp. 99–111.
5. Porter, J. P.: "A comparative study of some measures of persistence" (abstract). *Psychol. Bull.*, Vol. XXX, 1933, p. 664.
6. Ryans, D. G.: "An experimental attempt to analyze persistent behavior: II. A persistence test." *J. gen. Psychol.*, Vol. XIX, 1938, pp. 355–371.
7. Thornton, G. R.: "A factor analysis of tests designed to measure persistence." *Psychol. Monog.*, Vol. LI, No. 3, 1939.
8. Thornton, G. R.: "How general is the factor of 'persistence'? A re-examination and evaluation of Ryans' results." *J. gen. Psychol.*, Vol. XXI, July 1940, pp. 185–189.
9. Williamson, E. G.: *How to counsel students.* New York: McGraw-Hill, 1939.
10. Woodworth, R. S.: *Psychology.* Fourth Edition. New York: Henry Holt, 1940.

A STUDY OF FAILURES ON THE STANFORD BINET IN RELATION TO BEHAVIOR AND SCHOOL PROBLEMS

IRA S. WILE AND ROSE DAVIS

New York City

The intelligence quotient has become an index of potential educational function. Regardless of the relative significance of innate factors and environmental stimulation, it has gained respect and favor in the classification of children and in establishing prognoses concerning their probable school progress.

The IQ, being dependent upon an ascertained mental age, carries with it an implication of equality of functional growth at identical mental ages. This is contrary to the fact as mental ages differ functionally in terms of their technical constitution. Thus two children, *A* and *B*, each with a mental age of ten years, four months, may have tremendous variation in intellectual capacity. If *A* had a basal age of nine years while that of *B* was seven years, the fundamental developmental levels would be very dissimilar. Furthermore if *A* and *B* had the same basal age of seven years their cerebral endowment might still vary greatly, as *A* might have a spread of accomplished tests above the BA wholly dissimilar to those achieved by *B* at the higher age levels. The BA represents an equality of development for those at the same level while the MA contains possible inequality because of the different tests allocated and achieved above the BA and entering into the MA.

Assuming IQ to have significant index value, it cannot be interpreted intelligently without reference to the basal age and the specific tests passed or failed. From a clinical standpoint it is essential to understand the functional capacity of the examinee. This is tantamount to interpreting the psychological strength and weakness of mental function. Specific success or failure above the basal age helpfully indicates the existence or non-existence of a psychological trait at any arbitrarily chosen age level.

The student of behavior problems is concerned with the child and should seek to know all factors endogenous and exogenous that possibly enter into his behavior responses. The IQ, *per se*, is not a sound basis for explaining personal reactions, even though it may be a dominant factor in a number of undesired reactions. This theoretical index is but an expression of responsiveness to a definite test situation.

It forms one facet of personality, but in and by itself the IQ does not indicate the functional elements giving rise to its manifestation.

Children with very high IQ's may and do fail in scholastic work, while some with average or dull normal status may plod toward success. What are the reasons for their failure or success? If it cannot be explained by ascertained differences in personality traits such as industry, ambition, determination and the like, possibly the cause may be found in the mental organization of the child.

Emotional factors as immaturity, instability, inferiority, may militate against success on an psychological test, but repeated failure in scholastic work may conduce to many behaviors of maladjustment. Aggressiveness or submissiveness, worry and despair may be effects of inherent difficulties affecting the BA, MA and IQ, and in turn affect the interest in and application to scholastic work. This is manifested in various studies of the personality of children with reading disabilities.[1]

Studies of the relation of reading ability to intelligence usually fail to disclose why some children with IQ's of 120 and over are poor readers while others with IQ's of 80–90 are relatively excellent readers. Marked differences of power at various intelligence levels are to be expected in terms of averages, but specific psychological limitations may be of greater importance than the total power suggested by MA or IQ. Leavell and Sterling[2] are correct in stating that there is a definite tendency for the less intelligent children to make more visual regressions and more fixations, to have lower reading rate and comprehension scores and to have a narrower span of recognition than the more intelligent children. But reading rate and comprehension are results of the other factors noted and still others more basic in the reading process but not made evident in a gross consideration of intelligence.

As children learn to appreciate differences before likenesses, and acquire the ability to copy a square before they can a rhombus, it is certain that maturation involves a growing capacity to function along any particular line. The gain in power of aural recall marks maturation, but the fact that digits directly recalled may be many more in number than those recalled in reversed order indicates that these

---

[1] Challman, R. C.: "Personality Maladjustments and Remedial Reading." *J. Except. Child*, Vol. VI, 1939, pp. 7–11.

[2] Leavell, U. W., and Sterling, H. A.: "A Comparison of Basic Factors in Reading Patterns with Intelligence." *Peabody J. Educ.*, Vol. XVI, 1938, pp. 149–155.

processes of recall contain elements essentially different. Perhaps the ability to recall in direct given order is a manifestation of an immediate auditory memory of rote type without an accompanying retentive power of memory that permits recall in reverse of given order through a structural appreciation of the number series. There appears to be a failure in establishing an auditory configuration that gives meaning to the component parts in their relationship. If this be true, the psychological organization of the examinee is partially revealed by his successes and failures to serve for a partial interpretation of his general and specific responses.

In an effort to ascertain the psychological significance of the IQ, BA, and certain specific tests entering into the MA, we made a study of one hundred children who were referred to The Children's Health Class of Mt. Sinai Hospital because of such problems as aggressive and submissive behavior, tics, speech defects and enuresis. Routine examinations revealed the concurrence of other disorders of personality or behavior incidental to intersibling conflicts, infantile position in the family and difficult adjustment in the school situation. We call attention to the fact that our study involved some case selection because (1) these children had been brought to a clinic for the treatment of their behaviors and not because of any educational inadequacy; (2) mental defectives were excluded from the study; (3) the clinic itself does not receive a typical cross section of the community. Bearing these restrictive factors in mind, the problems of the one hundred children may be tabulated as follows:

PROBLEMS PRESENTED BY 100 CHILDREN

	PER CENT
Problems referred	
Aggressive behavior	50
Temper tantra, pugnacity, disobedience, etc.	
Submissive behavior	20
Phantasy life, retiring unhappiness, no friends, etc.	
Tics, habits of nail biting, thumb sucking	40
Speech difficulties	36
Enuresis	23
Problems noted	
Sibling conflicts	20
Infantile position in the home	20
Difficult adjustment in school	75

The striking prevalence of school maladjustment suggested that the various problems might have some common psychological denom-

inator. We, therefore, instituted a study of their failures on the Stanford Revision of the Binet-Simon Tests and found pronounced innate difficulties that illuminated all the problems to a degree unattainable by mere reference to the IQ or the Mental Age.

The children can be presented statistically as follows:

		Per Cent
CA range from five years to twelve years	5–6¹¹ years	12
	7–9¹¹ years	72
	10–11¹¹ years	16
MA (Binet) from three years to twelve years	3–6¹¹ years	12
	7–9¹¹ years	82
	10–11¹¹ years	6
IQ (Binet) from 80 to 130	80–90	24
	90–99	34
	100–109	30
	110–130	12

No comparison of these figures is possible as they are grouped and not matched. It is clear, however, that considering the sixteen per cent with CA 10–11¹¹, the MA 10–11¹¹ were far less numerous (six per cent). The IQ distribution would be skewed naturally as all children with IQ's below 80 were rejected so as to lessen the problem of psychological interpretation by assuming some essential deficiency in cerebral structure. It is notable that the IQ 110–130 group varies but little from expectancy while the IQ 80–90 group is larger than would ordinarily be expected.

It is not possible to attribute the seventy-five per cent of school maladjustments to low IQ when sixty-four per cent of the children had IQ's of 90 to 100 and twelve per cent were above IQ 110. Hence, it is important to investigate the basal ages of the various MA groups. These were as follows:

12 Per Cent
MA 3–6¹¹, Range IQ 86–112

		Age	Number
Basal age range 3–6 years	BA	3	5
	BA	4	0
	BA	5	4
	BA	6	3

82 Per Cent
MA 7–9^{11}, Range IQ 80–130

		Age	Number
Basal age range 5–8 years	BA	5	5
	BA	6	45
	BA	7	6
	BA	8	3
	Two BA	6, 8	29
	BA	9	0

6 Per Cent
MA 10–11^{11}, Range IQ 90–113

		Age	Number
Basal age range 6–8 years	Two BA	6, 8	3
	BA	8	3

A glance at this distribution reveals the large number of relatively low BA's in each MA group. This becomes increasingly evident as one proceeds to each higher MA group noting the highest MA and highest BA.

Even more striking is the increasing proportion of double basal ages which reveals a definite weakness in functional development, *i.e.*, some factor has not evolved so as to permit an orderly progression of the MA above the structuralized function indicated by the basal age. It seemed necessary to ascertain the reason for the double BA presented by twenty-nine children out of eighty-two (thirty-five per cent) with MA 7–9^{11}, and by three children out of six (fifty per cent) with MA 10–11^{11}. We, therefore, made a careful review of all the replies to the tests with especial attention to all failures.

The results of the investigation were very definite and cast much light upon the reasons for educational difficulties and also for the secondary behavior responses incidental to emotional factors stemming from an actual psychological inadequacy. The failures of the various age groups naturally varied somewhat because of the differing demands of the tests at various ages. There is nonetheless most supportive evidence of very definite lines of psychological non-development. The statistics of failures on certain tests clarify the nature of the psychological deficiencies, without reference to the IQ.

GROUP *A* (12 PER CENT)

MA 3–6^{11} YEARS	NUMBER
Failure to draw a square and recognize designs (IV)	8
Failure to distinguish left and right (VI)	5
Failure to draw diamond (VII)	9
Failure to repeat digits (IV)	4
Failure to repeat digits in reversed order (VII)	8

GROUP *B* (82 PER CENT)

MA 7–9^{11}	NUMBER
Failure to draw diamond (VII)	74
Failure to repeat digits in reversed order (VII)	68
Failure to repeat digits (VII)	30
Failure to distinguish left and right (VI)	5
Failure with vocabulary (oral) (VIII)	22
Failure in design reproduction (X)	20

GROUP *C* (6 PER CENT)

MA 10–12 YEARS	NUMBER
Failure to draw diamond (VII)	3
Failure to repeat digits in reversed order (VII) (IX)	6
Failure to repeat digits (X)	4
Failure with vocabulary (X)	4
Failure in design reproduction (X)	5
Failure in reading test (X)	6

That eight children of the twelve (Group *A*) failed to draw the square; nine of twelve (Group *B*) and three of six (Group *C*) failed to draw the diamond is most significant and suggests a limitation in visual learning, observation, memory and recall plus a kinesthetic factor. There is an implication in this fact that eye and hand dominance may be of importance and serve as one reason for Dearborn's[1] emphasis upon the left ocular or manual dominance so frequently found among retarded readers.

The limitation in mental representation is not only in the visual field, as is shown by the fact that eight of twelve (*A*), sixty-eight of eighty-two (*B*) and six of six (*C*) were unable to repeat digits in reversed order. This suggests a lack of capacity in auditory memory, mental representation and recall plus a possible kinesthetic articulatory factor. Larson and Feder[2] report that "Reading memory, hearing memory, complex sound discrimination, and phonics did not differ-

[1] Dearborn, W. F.: *The Nature and Causation of Disabilities in Reading.* Suppl. Educ. Monog. Vol. IL, 1939, pp. 103–110.

[2] Larson, Robert P. and Feder, D. D.: "Common and Differential Factors in Reading and Hearing Comprehension." *Jour. of Educ. Psychology,* Vol. XXXI, April, 1940, pp. 250–251.

entiate sufficiently to be considered significant conditioners of skill in reading and hearing comprehension. Oral reading accuracy was the only differential factor showing a fair degree of relationship to reading comprehension, and a slight negative relationship to hearing comprehension. Comprehension must depend in part upon the dynamics of learning, and these include auditory as well as visual responsiveness and an attendant capacity to recall and associate form and sound. A lack of this psychological factor may affect the BA, MA, IQ and scholastic aptitude and achievement.

That some factor of mental confusion was operative is suggested by the fact that five of twelve (*A*), five of eighty-two (*B*) could not distinguish right and left; also, five of six (*C*) could not reproduce the designs (X). Most important of all is the inability of the six (*C*) with MA's 10–12 to pass the ten-year reading test while four of them were unable to pass the vocabulary test allocated to their chronological age. The numbers are not large and apparently the IQ's and MA's do not suggest the presence of mental confusion, so it was decided to study results from the Kohs Block test. We have shown elsewhere[1] its value and find that it offers significant information through an intelligence quotient based upon performances involving form, color, and position. The manipulation of color pattern requires clear observation plus a slight amount of mental representation with recall.

The Kohs Block Test showed an IQ range of 60–130. The distribution of IQ's was in sharp contrast with that on the Binet test.

IQ range	Koh's per cent	Binet per cent
60–79	45	0
80–89	20	24
90–99	16	34
100–109	13	30
110–130	6	12

It is now patent that the Kohs Block Test revealed a weakness, a possible confusion, at least an incompetency which is in harmony with the types of failures in the visual field as tested on the Binet. Whereas we had excluded all children with IQ's of 80 or less, we now found

[1] Wile, Ira S. and Davis, Rose: "The Use of the Kohs Test as an Indicator of Mental Confusion." *The Am. Jour. of Orthopsychiatry*, Vol. VI, No. 1, Jan., 1936, pp. 1–16.

forty-five per cent registering IQ's (Kohs) below 80, sixty-five per cent below 90, and only six per cent with IQ's of 110 and above. Seemingly the seventy-five with school maladjustment were more readily revealed in the sixty-five per cent with IQ's (Kohs) below 90. Inadequacies in visual percepts and memory, with pattern reversals, supplement poor visual span and at least hint that the visual inadequacies for form and design, recognition, copying and reproduction, play a definite part in causing confusion in the learning process. This does not argue that the psychological weakness is due only to mixed eye-hand function, nor to the effects of conversion from dominant handedness. It lends some support to the idea that sound discrimination and auditory memory are not the basic factors in handicapping learning or in causing reading disabilities.

The school difficulties of our group attest the probable correctness of the interpretation that a specific psychological weakness in visual memory and visual auditory association underlines the failure of most of these children in school and non-school adjustment. Their school grades ranged from the kindergarten to 6B, inclusive.

	Per Cent
Correct grade placement for CA	63
Grade retardation from one to four terms	37
Marked reading and spelling disability	35
(with grade range from 1B to 6B for IQ range 80–120)	
Marked arithmetic disability	15
(with grade range from 2B to 6B for IQ range 85–115)	
Scholastic proficiency in reading and arithmetic below the standard of grade placement	38

While visual and aural associations are essential factors in learning, especially in acquiring reading facility, mental confusion may be added by interference with natural cerebral dominance. This causal factor cannot be neglected in endeavoring to correct the fundamental psychological weakness in the visual-auditory field. This is suggested by the difficulty of some of the younger children in differentiating right and left and by the confusions shown in solving the forms in the Kohs Block Test where many reversals were manifest.

Without analyzing the frequency of strephosymbolia or the peculiarities of oral reading and spelling, it is interesting to find that thirty-three per cent of the group were converted sinistrals (grade range Ktgn. to 6B; IQ 80–130).

Fifty-nine per cent had mixed eye-hand function

Thirteen per cent right eyed and left handed
Forty-six per cent left eyed and using right hand.

The importance of this confusing status is shown in the fact that seventy-two of the groups revealed definite reversals, transpositions and confusions in writing letters or numbers, in oral and written spelling, in reading aloud, solving problems or performing computations.

The poor readers in our groups were not brought to the clinic because of their maladjustment. Their IQ's alone would not have revealed their potentials for school failure. The BA gave a slight indication that psychological development was not adequate or regular despite a high MA. The analysis of failures and the results upon the Kohs Block Test demonstrated that the children were sufferers from an incapacity in perceiving, memorizing, recalling and associating symbols in line or sound. They were, therefore, weak in the power of acquiring verbal or symbolic material through the eye. They were failures in school and, possibly, in reasonable emotional responses because they could not master the mechanics of reading and arithmetic.

While symbol scales[1] have been devised to test the naming and mastery of letters on the basis of giving their sound, printing and writing letters and making phonic combinations, it is certain that the fundamental weakness can be observed in the course of routine psychological testing. The extent of the weakness and the particular emphases or methods essential for correcting the difficulty can be readily determined by achievement tests or by the more extensive tests devised for studying the factors entering into reading disabilities.

In general it may be said that remedial work should not be restricted to practical reading, or arithmetic, or to drills upon phonetic or flash methods of approach. The attack should be made upon the various psychological shortcomings in the field of visual-aural-kinesthetic associations, together with the rectification of any confusions incidental to a lack of unilateral cerebral dominance. The comprehension skills are founded upon skills in the mechanics of learning and these must depend upon psychological abilities, as well as upon total mental capacity and organization of the personality.

BIBLIOGRAPHY

1. Challman, R. C.: "Personality Maladjustments and Remedial Reading." *Journal of Exceptional Children*, Vol. VI, 1939, pp. 7–11.

[1] Wilson, F. T. and Fleming, C. W.: "Symbol Scales for Use in Beginning Reading," J. Psychol., Vol. VIII, 1939, pp. 99–114.

2. Dearborn, W. F.: *The Nature and Causation of Disabilities in Reading.* Supplemental Educational Monograph, Vol. IL, 1939, pp. 103–110.
3. Larson, Robert P. and Feder, D. D.: "Common and Differential Factors in Reading and Hearing Comprehension." *Journal of Educational Psychology,* Vol. XXXI, No. 5, 1940, pp. 250–251.
4. Leavell, U. W. and Sterling, H. A.: "A Comparison of Basic Factors in Reading Patterns with Intelligence." *Peabody Journal Education,* Vol. XVI, 1938, pp. 149–155.
5. Wile, Ira S. and Davis, Rose: "The Use of the Kohs Test as an Indicator of Mental Confusion." *The American Journal of Orthopsychiatry,* Vol. VI, No. 1, 1936, pp. 1–16.
6. Wilson, F. T. and Fleming, C. W.: "Symbol Scales for Use in Beginning Reading." *Journal Psychology,* Vol. VIII, 1939, pp. 99–114.

AN EXPERIMENTAL ANALYSIS OF THE CONTRIBUTIONS OF SPEED AND LEVEL IN AN INTELLIGENCE TEST

BRENT BAXTER

University of Minnesota*

Speed and level, as factors of intelligence, have been the subject of many experimental studies. The earliest approach to the study of the relationship of these factors was to correlate the scores obtained by a group in standard or limited time (speed) on an intelligence test with the scores obtained in an unlimited time (level). This method revealed that under most conditions the correlation was fairly high, often in the .90's. With this evidence it was concluded that only a few students were handicapped by limited time and that these two measures of intelligence were much the same. The limited time score, however, is not the only measure of speed or rate obtainable from testing procedures. It is also possible to measure the amount of time to finish the entire test. As a measure of rate the latter is to be preferred, for all subjects cover the same material and can be compared in terms of time for which there are equal units. In the limited time score the subjects cover different amounts of work, and the score is in terms of units of unequal or unknown difficulty. It appears, therefore, that the "time to finish" score is the best measure of the factor called speed.

In order to have a measure of speed that would be as free as possible from level, some studies[3,5] have measured speed over very simple mental tasks such as easy arithmetic or sentence completion items. When this measure was correlated with standard (limited) time scores on intelligence tests, the average correlation was about .30, indicating that one's speed of doing simple items is only slightly related to the number of correct answers for difficult material in a limited amount of time. It may be pointed out, however, that actually both of these scores are measures of rate and that, therefore, the relation between speed and level is obscured. Raskin[4] was interested in noting how the relationship would change if speed were measured over more difficult material. She concluded that "speed is, to a large extent, linked to the difficulty of the material on which it is measured."

---

* Psychology Laboratory of the University of Minnesota, under the direction of Dr. Miles A. Tinker.

The problem of the present study was derived from the questions raised by this earlier research. The study aimed to secure separate measures of speed and level over items from all levels of difficulty. Moreover, it sought to note the relationships of speed and level in the scores obtained in limited time. Further, it wished to discover their validity in terms of "outside" criteria of intelligence.

The Otis Self-Administering Test, Higher Examination, Form *A*, was administered individually with special directions to one hundred University of Minnesota sophomores. The subjects were instructed to work for both speed and accuracy, taking each question in turn without going back to any items already attempted. At the end of twenty minutes they were given a different colored pencil and told to complete the entire seventy-five items. The experimenter kept an account of the total time spent in completing the entire test. Three scores are obtainable from this procedure: (1) The number of items correct in limited time, (2) the number of items correct in unlimited time, and (3) the number of seconds required to attempt the entire test. The scores are redefined in a way that is more meaningful in terms of the data to be presented. The first score instead of being speed is called power, the second is called level, and the third is speed.

Three outside criteria of intelligence were selected. The Army Alpha Revised (Wells) Form Five was given under shortened time limits (sixteen minutes instead of twenty-one and one-half minutes) during the regular psychology laboratory period. The subjects had been given the Minnesota College Aptitude Test, one hundred vocabulary items, upon entering college. Their honor point ratio was based upon the first forty-five credits (or first year's study) at the University of Minnesota; only academic credits were counted, all physical training and R.O.T.C. credits being omitted.

RESULTS

Our first interest is in the relations between the three scores obtained from the Otis test. The correlations and their probable errors are:

$$\text{speed } vs. \text{ power} = .75 \pm .03$$
$$\text{speed } vs. \text{ level} = -.06 \pm .07$$
$$\text{power } vs. \text{ level} = .62 \pm .04$$

The most notable correlation here is that between speed and level. Because of its large probable error, the correlation of $-.06$ may be said

to be not significantly different from zero. The measures of speed and level vary independently; the length of time that the student works has no relation to the quality of that work. This fact has been reported in other studies.[1] The correlations between power and level and between power and speed represent fairly high relationships. However, as pointed out before, these scores are bound to have a sizable correlation because of their common relations in the testing procedure.

One of the aims of this study was to find the relations of speed and level in the power score. The multiple correlation of speed and level with power is .998, the coefficient of multiple determination being .996 ($.998^2$). Similarly, if the zero order coefficients of determination are added (this may be done since speed and level vary independently), the sum is .95.* If the beta coefficients are squared and added, the sum is 1.06.† These results agree in indicating that speed and level are the unique factors which are determining power. This is entirely reasonable. We may make an analogy of these scores with the climbing of a difficult mountain. One's position on the mountain after a period of climbing (power) is entirely dependent on one's skill in climbing (level) and the rate at which one climbs (speed).

Which of these two factors, speed and level, is the more important in determining the limited time score? If the zero order coefficients of determination (.56 for speed and .39 for level) are compared, it indicates that speed determines about sixty per cent of the total variance of the power score. If the squares of the beta coefficients (.62 for speed and .44 for level) are compared, the same answer is obtained. Thus it may be said that the score that the individual obtains in standard or limited time is dependent upon his rate of work to an extent of sixty per cent. Also, since it has been shown that speed and level contribute the entire variance of power, level determines the remaining forty per cent of the score in standard time. It must be remembered, of course, that these results hold only for the Otis test and for the particular limited time used in this experiment. It would be interesting to know how the relations between speed and level vary throughout the test. If the power score were obtained at the end of five minutes in which the relatively easy items were attempted, it might be expected

---

* $.75^2$ plus $.62^2$ equals .56 plus .39 equals .95.

† The beta coefficient for speed *vs.* power and level is .66; for level *vs.* power and speed, it is .79.

that speed would be even more important in determining it than at the end of twenty minutes which would include more difficult items.

Table I presents the correlations of the Otis test scores and the outside criteria of intelligence, Army Alpha Revised (Wells) Form Five, honor point ratio, and the Minnesota College Aptitude Test.

TABLE I.—CORRELATIONS BETWEEN SPEED, POWER, AND LEVEL, AS MEASURED ON THE OTIS SELF ADMINISTERING TEST AND THREE CRITERIA OF INTELLIGENCE

	Honor point ratio	College Aptitude Test	Army Alpha Test
Speed	.28	.18	.52
Level	.47	.44	.50
Power	.49	.43	.68

First we shall consider the relationships between the Otis variables and honor point ratio. Regarding the zero order correlations, the power score has the greatest validity, and level is only very slightly lower. Speed, however, is considerably lower than either of the other variables in validity, as noted if .28 is compared with .49 and .47. The multiple correlation between honor point ratio and speed with level is .56. This means that speed and level account for .31 ($.56^2$) of the variance of honor point ratio. We are interested in noting the relationships of speed and level in this total variance. This may be found by examining the beta coefficients again. Below is shown the steps by which the relationships are found.

$$B_{(HPR \text{ and } S)} \cdot L = .31 \qquad (.31^2) = .09 \qquad \frac{.09}{.33} = .29$$

$$B_{(HPR \text{ and } L)} \cdot S = .49 \qquad (.49^2) = .24 \qquad \frac{.24}{.33} = .71$$

$$\text{Sum of the betas squared} = .33$$

The beta coefficients for speed and level have been squared and then summed. This total, .33, represents the total variance contributed to honor point ratio by speed and level and compares with the square of the multiple correlation coefficient or .31. The percentage of this total contributed by each variable was determined; speed contributes twenty-nine per cent and level contributes seventy-one per cent. In

comparing the valid (as determined by honor point ratio) contributions of speed and level, it may be said that level is about two and one-half times as valid as speed. In other words, the scores obtained in unlimited time are much more useful in determining college grades than are the times required to finish the test.

The correlation between honor point ratio and power is .49. The variance that power contributes to honor point ratio is .23 ($.49^2$). Previously it was found that level contributed forty per cent of the variance of the power score. However, it cannot now be said that level contributes forty per cent of the variance that power contributes to honor point ratio. It has just been found that level is more valid as a measure of honor point ratio than speed, and thus probably makes a greater contribution to the validity of the power score. In order to find the relative size of the contributions of speed and level to the valid variance of power, a modified method of partial correlation proposed by Franzen[2] will be used. In the case of three variables, the correlation is written $r(1 \cdot 3)2$, which is read as the correlation between factors 1 and 2, with the effects of factor 3 removed from 1 but not from 2. If Franzen's method does what he claims and is valid, this is just the type of correlation that is wanted to answer our question. We first must refer to the evidence previously presented; that speed and level contribute one hundred per cent of the variance of power. The problem now is to note the relationship between honor point ratio and power with first the effects of speed and then level removed. If either level or speed is removed from power, the other variable is left. The steps in computing our answers are shown below.

r (power · speed) HPR	$= .42$	$.42^2 = .18$
r (power · level) HPR	$= .25$	$.25^2 = .07$
Total variance of speed and level in valid power		$= .25$

The partial correlations were squared so that we are dealing with variance. The total variance here may be compared with the total variance that was found by squaring the correlation between honor point ratio and power, which was .24. If each of the contributions of speed and level is related to the total variance, it is found that level contributes seventy-two per cent, speed contributing the remainder, twenty-eight per cent. Though level contributes only about two-thirds (40 divided by 60) as much as speed to the variance of power,

since level is two and one-half times as valid, level contributes two and six-tenths times as much as speed to the valid variance of power. In other words, whatever is determining the scores in unlimited time is not of major importance in determining the score in limited time, but it is comparatively very important in determining that "aspect" of the limited score that is useful in predicting college grades.

The same calculations may be performed with the Army Alpha test as the criterion of intelligence. The multiple correlation of speed and level with Army Alpha is .74. Of their total valid variance ($.74^2$ or .55), speed contributes fifty-one per cent and level the remaining forty-nine per cent. Of the valid variance of power ($.68^2$ or .46), speed contributes fifty-three per cent and level forty-seven per cent. Thus, the unlimited time scores and the times required to finish the test are about equally efficient in predicting the Army Alpha scores and are also about equally associated with that "aspect" of the limited time scores which is associated with the Army Alpha scores.

Similarly with the college aptitude test, the multiple correlation with speed and level is .49. Of their total valid variance, speed contributes seventeen per cent, and level the remaining eighty-three per cent. These same percentages also indicate the relative contributions of speed and level to the valid variance of power. Thus, the unlimited time score is almost five times as efficient as the time to finish score in determining the scores on the vocabulary test. The unlimited time score is equally superior in determining that "aspect" of the limited time score that is associated with the scores on the vocabulary test.

The results concerning the outside criteria are summarized in Table II.

TABLE II.—RELATIVE PERCENTAGES OF CONTRIBUTION OF SPEED AND LEVEL OF THE OTIS TEST AS DETERMINED BY THREE CRITERIA OF INTELLIGENCE

	Honor point ratio	Army Alpha Test	College Aptitude Test
Speed	29	51	17
Level	71	49	83
Speed in power	28	53	17
Level in power	72	47	83

Table II emphasizes two points. (1) The relationships of speed and level to the criteria are about the same as their relationships in determining the valid (according to the same criteria) variance of the

power score. (2) The relationships of speed and level vary with the criterion of intelligence, level being most valid when the college aptitude test is used, and least valid when Army Alpha is used. That speed is a much more important factor with Army Alpha than either of the other criteria might have been expected *a priori*. Army Alpha is considered a "speed" test at the college level since almost all of the items are within the grasp of college students and the test is administered under severe time limits. On the other hand, a vocabulary test, such as the college aptitude test, has always been considered an excellent measure of level, and here it is shown that level, as we measure it, does play a very important rôle in the variance of the test scores.

It is also interesting to note that speed correlates more highly with honor point ratio than with the college aptitude test. If our speed score is a measure of work habits this same factor is somewhat more important in determining school marks than scores on vocabulary tests. This may be due to the trend toward giving school marks on the basis of objective examinations given in limited periods.

GROUP TESTING RESULTS

Form *B* of the same test was administered to the same individuals in groups of between thirty and forty students one month, on the average, after the original testing. This was done to see if the relationships involved in the factors would change under the conditions of group testing. Also in this manner some estimate of the reliability of the scores would be made available.

Otis reports a reliability coefficient of .92, Form *A vs.* Form *B*, for one hundred and twenty cases, grades IX through XII. This, of course, would refer to the standard time score, here called power. The reliabilities found in this experiment are shown in Table III.

TABLE III.—RELIABILITY COEFFICIENTS OF THE OTIS TEST VARIABLES

Speed (Form *A versus* Form *B*)	.63
Power (Form *A versus* Form *B*)	.70
Level (Form *A versus* Form *B*)	.72
Power (odd *versus* even, Form *A*)	.93

The test-retest coefficients in Table III are of approximately the same magnitude but are significantly lower than that reported by Otis. Considering the conditions under which the reliability was obtained, however, these coefficients are not out of line with what might have been expected. Form *A* was given individually, but Form *B* was

given in groups. Also the retest was administered at varying intervals, ranging from one week to two months. Furthermore, our group is highly homogeneous as compared to Otis' group. The odd-even coefficient for the power score of Form *A* indicates a high internal reliability. It is quite possible, then, that a score on either of the Forms is highly reliable itself, but that it indicates the student's rank only for a single set of conditions. Further research is needed to determine this definitely.

The correlations indicating the relationships between the Otis scores for Form *B* and group testing are:

$$\text{speed } vs. \text{ power} = .73 \pm .03$$
$$\text{speed } vs. \text{ level} = .17 \pm .07$$
$$\text{power } vs. \text{ level} = .71 \pm .03$$

These correlations are not significantly different from those obtained for Form *A*. Again speed and level vary independently while both of them show an overlap with power. Using the same analysis as with Form *A*, it was found here that speed contributes fifty-three per cent of the variance of the power score which compares with the sixty per cent found on Form *A*. This difference is probably not significant since the correlations on which the percentages are based did not differ significantly.

The next step is to analyze the results to see if the validity of the variables changes from one type of testing situation to the other. Table IV presents the zero order correlations between the Form *B*

TABLE IV.—ZERO ORDER CORRELATIONS BETWEEN THE OTIS VARIABLES OF FORMS *A* AND *B*, AND THE CRITERIA OF INTELLIGENCE

	Honor point ratio	Army Alpha Test	College Aptitude Test
Form *A* speed	.28	.52	.18
Form *B* speed	.28	.60	.05
Form *A* power	.49	.68	.43
Form *B* power	.40	.63	.39
Form *A* level	.47	.50	.44
Form *B* level	.28	.44	.34

Otis variables and the criteria of intelligence. It also presents the correlations for Form *A* for purposes of comparison.

None of the changes in correlation from Form *A* to Form *B* are statistically significant at the one per cent level of probability. The only change that approaches it (three per cent level) is for the decrease in validity of the level score as determined by honor point ratio. There is a trend for the level score to decrease in validity according to all three of the criteria. Accompanying this decrease in the validity of the level score is a decrease in the validity of the power score. This is as might be expected, since the power score is to some extent dependent upon level.

The procedure for analysis of Form *A* was again applied to Form *B*. Table V presents the summary, indicating the relative valid contributions of speed and level. This table is comparable to Table II.

TABLE V.—RELATIVE PERCENTAGES OF THE CONTRIBUTION OF SPEED AND LEVEL OF THE OTIS TEST (FORM *B*) AS DETERMINED BY THREE CRITERIA OF INTELLIGENCE

	Honor point ratio	Army Alpha Test	College Aptitude Test
Speed	50	65	1
Level	50	35	99
Speed in power	50	68	14
Level in power	50	32	86

While the percentages in Table V are different from those in Table II, they probably do not differ significantly, for they are based on zero order correlations which are not significantly different. The percentages of Table V only bear out the trend in the decrease in the validity of level under the conditions of individual testing.

EFFECT OF RESULTS ON TESTING PROCEDURES

What relation do these results have to the practical testing procedures? Some of the questions to be answered are: (1) Is the present method of using the score in standard time the most valid single score obtainable (as compared to our other measures)? (2) Is speed a measure worth considering in testing? (3) Do speed and level measured separately improve on prediction as compared to the power score?

The zero order correlations of speed, power, and level with the criteria of intelligence (Table IV) will show the relative validity of our measures. In relation to honor point ratio, power has greater validity

than either speed or level, though level has about equal validity. With Army Alpha, power is far superior in validity. For College Aptitude Test, level has a greater validity though here again power has approximately equal validity. Thus, we may say in general that power has superior or equal validity as compared to speed or level. Is there any *a priori* reason for suspecting that this would be the case? We found that both speed and level have a certain validity, and so it seems reasonable to expect that some combination of these two variables would be even more valid than one separate measure. This combination has been shown to exist in power. Recently there has been a tendency to give students a very long period of time to complete all college aptitude tests. These results suggest that at least equally valid scores might be obtained by administering the tests in shorter periods. Considerable time might thus be saved without any loss in prediction efficiency.

Next we must answer the question as to whether or not this combination of speed and level in the power score has greater validity than their combination in multiple correlation. Honor point ratio correlates to an extent of .49 with power, with speed · level, .56; Army Alpha and power correlate .68, with speed · level, .74; College Aptitude Test correlates .43 with power, with speed · level .49. In every case the multiple correlation of speed and level correlates higher with the criterion than the power score, the amount varying from six to ten points. Are these differences significant? If the "z" test is applied to these differences, we find that all of them are highly significant, none of them occurring by chance more than once in one hundred times.

It has been shown, then, that by taking separate measures of speed and level and using them in a multiple regression correlation, an increase in validity from six to ten points over the usual score in standard time is obtainable. This increase in the amount of correlation will increase our predictability in varying amounts, depending upon the range of correlation in which the increase is made. With honor point ratio a four and one-half per cent improvement in prediction is made by combining speed and level in multiple correlation. For Army Alpha and College Aptitude Test the improvement is six and three and one-tenth per cent, respectively. Thus, it would seem advisable, especially in the case of honor point ratio, in view of prediction alone, to obtain measures of speed and level separately. These improvements in prediction are based on the variables of the Otis as measured by individual testing. In the case of group testing, however, we noted

a tendency for the validity of level to decrease, which corresponded with a decrease in the validity of power. Because of these decreases, any prediction on the basis of Form *B* results is considerably less than the prediction from the individual testing results. The combination of speed and level in multiple correlation is no longer superior to power. The improvement in prediction seems to be dependent upon the careful individual administration of the test. Therefore, in the practical situation we must weigh against the improvement in prediction the large increase in testing time. For testing programs in which a large number of subjects are to be measured, it seems inadvisable to obtain separate scores. For individual cases and small groups, however, the increase in predictability seems well worth the added time and effort.

SUMMARY AND CONCLUSIONS

An analysis of the relationships of speed, level, and power as measured on the Otis Self-Administering Test, Higher Examination, was made. The factors were also compared with three outside criteria of intelligence: Honor point ratio, Army Alpha Revised (Wells) Form Five, and a college aptitude test. Form *A* was administered individually to one hundred college sophomores and Form *B* was given to the same subjects in groups. In both cases the subjects were instructed to work for both speed and accuracy. From the test three scores were made available and defined as follows: (1) The number of seconds to complete the entire seventy-five items (speed), (2) the number of items correct at the end of twenty minutes (power), (3) the number of items correct out of the seventy-five (level).

The most important conclusions concerning the relationships of speed and level in intelligence testing disclosed by the results are:

(1) Speed and level, as measured in this investigation, vary independently.

(2) Speed and level contribute the entire variance of power, or score in standard time, speed contributing slightly more than level.

(3) The valid relationships of speed, level, and power vary with the criterion of intelligence.

(4) The relationships of speed and level in the criteria are about the same as their relationships in accounting for the valid variance of power.

(5) There is a tendency for the validity of level and power to decrease when measured in groups as compared to individual testing.

(6) Prediction through the combination of speed and level in multiple correlation is greater than that possible through the use of scores in standard time, though the tests must be given to each individual separately.

REFERENCES

(1) Freeman, F. S.: "The factor of speed." *J. Gen. Psychol.*, Vol. VI, 1932, pp 462–468.

(2) Franzen, R.: "A comment on partial correlation." *J. Educ. Psychol.*, Vol. XIX, 1928, pp. 194–197.

(3) Hunsicker, L. M.: *A Study of the Relationship between rate and ability.* Teachers Coll. Publ., Columbia Univ., No. 185, 1925, pp. 52.

(4) Raskin, E.: *The interrelationships of speed of response measured on levels of varying difficulty.* A Ph.D. thesis on file in the Library of the University of Minnesota, 1937.

(5) Walters, F. C.: *A statistical study of the time factor in intelligence.* Teachers Coll. Publ., Columbia Univ., No. 248, 1927, pp. 84.

NOTES ON THE RAPID CALCULATION OF ITEM VALIDITIES

ALPHONSE CHAPANIS

Yale University*

Mental testers seem fairly well agreed that the biserial coefficient of correlation† is one of the best statistical means of estimating item validities.[5,7,10,12] The complexity of the operations involved and the time required to compute the biserial coefficient have, however, restricted its use.[1,12] Statements such as this are common: "The time consuming labor of 'best' existing methods makes the statistical selection of items impractical for those who should use it most, *i.e.*, the public school teachers."[1] It is the purpose of this paper to call attention to several short cuts which make the accurate determination of item validities by this means no more tedious or involved than some of the "time-saving" methods now in use.

The formula for the biserial coefficient of correlation may be written:

$$r_{bis} = \frac{M_p - M_f}{\sigma_T} \cdot \frac{pq}{z} \qquad (A)$$

$$= \frac{M_T - M_f}{\sigma_T} \cdot \frac{q}{z} \qquad (B)$$

in which:

M_T = the mean criterion score for all individuals taking the test.

M_p = the mean criterion score for individuals passing the item.

M_f = the mean criterion score for individuals failing the item.

σ_T = the standard deviation of the total distribution of criterion scores.

---

* This paper was prepared while the author was on the staff of the Williamson County Child Guidance Study, a Rockefeller Foundation project being carried out in Tennessee with the coöperation of the Tennessee Department of Public Health. The gist of the article evolved from considerations arising in connection with the analysis of a mass of test data collected in the study.

† This method of correlation is applicable to test data in which the criterion scores form a continuous and quantitative variable with the other variable, the item score, presented in a dichotomous classification. The procedures discussed in this paper refer only to this type of test data.

p = the proportion of individuals passing the item.
$q = 1 - p$ = the proportion of individuals failing the item.
z = the height of the ordinate of the unit normal curve which divides p from q per cent of the area under the curve.

The following discussion centers around formula (*B*) although the same argument, with slight modifications, holds for (*A*). As will be indicated later, formula (*B*) involves fewer computations than (*A*) and, for our purposes, it is somewhat shorter than a third variant which makes use of the difference between the mean criterion score of individuals passing the item and the mean criterion score for the group, $(M_p - M_T)$. In other words, formula (*B*) calls for the difference between the mean criterion score of all individuals taking the test and the mean score of individuals failing the item. This difference is reduced to a standard form by expressing it in units of the standard deviation, the standard difference then being weighted for the number of individuals failing the item.

In most practical situations, the tester is concerned with securing comparable validity values for the items of a test and is not interested in comparing the values obtained with those found for other tests. The significance of this is that the reduction of the difference in mean criterion scores to a standard form is unnecessary, or, in terms of formula (*B*), σ_T may be omitted. This modification results in what we may term the reduced biserial coefficient of correlation, which is written:

$$_rr_{bis} = (M_T - M_f)\left(\frac{q}{z}\right) \qquad (C)$$

The coefficients which this reduced formula yields are in every sense as accurate as those derived from the conventional formula; they are, in fact, proportional to the latter and may be readily converted into the ordinary biserial coefficients merely by dividing them by the standard deviation of the total distribution of scores.

Two further short-cuts may be applied to the calculation of the mean values required in formula (*C*): First, all calculations may be performed in units of the step-interval and no correction need be applied for the size of the interval, since it has a constant numerical value for any problem; and, second, if deviations are taken from the same step in every case, *i.e.*, the assumed mean is always taken at

the same interval, then the means called for in formula (*C*) may be replaced by the deviations of the true means from the guessed mean. The difference between the guessed mean and the scale zero is a constant and is cancelled in the subtraction $M_T - M_f$. As in machine computation methods, minus deviations are obviated by taking the zero deviation at the lowest interval in which any frequencies are recorded.

The final reduced biserial coefficient of correlation formula may now be written:

$$r_{bis} = \left(\frac{\Sigma f_T d}{N_T} - \frac{\Sigma f_f d}{N_f}\right)\frac{q}{z} \qquad (D)$$

Deviations are in step-units and the only necessary condition for the proper use of the formula is that the deviations for the distributions of failing scores on each item be taken from the same guessed mean. It should be noted that to return to the conventional biserial coefficient at this point it is necessary to multiply by i, the size of the step-interval, and divide by σ_T.

The reason for employing formula (*B*) rather than (*A*) should now be apparent. The mean of the total distribution of scores is constant for all item validities of the test so that it needs to be computed only once. This factor, even though constant, cannot be eliminated from the formula, however, because it is multiplied by a variable factor, $\frac{q}{z}$, and another variable quantity, $\left(\frac{\Sigma f_f d}{N_f}\right)\frac{q}{z}$, is subtracted from the product. Formula (*B*) is shorter than the third alternative which is based on the difference between the mean criterion score of individuals passing the item and the mean score of the total group, since, in general, individuals failing the item will have lower criterion scores and will coincide with the smaller deviation values.

That formula (*D*) is quite simple may be demonstrated by actual examples. All of the calculations necessary for the determination of validities on three items are shown in Table I. The class intervals of scores appear in column 1; column 2 shows the total distribution of scores earned on the test by one hundred ninety-eight children. This distribution is necessary to obtain the $\frac{\Sigma f_T d}{N_T}$ quantity. In column 3 are deviations in interval units from the guessed mean taken at the midpoint of the interval 20–24. This column of deviations is used unchanged for every item. Distributions of scores made by indi-

viduals failing Items 1, 2, and 3 appear in columns 5, 7, and 9. The obtainment of these distributions constitutes the major part of the work, but systematic tabulating by methods similar to those suggested by Dunlap[4] will reduce this to a fairly quick routine step. In columns

TABLE I.—ILLUSTRATING THE COMPUTATION OF ITEM VALIDITIES FOR THREE ITEMS BY MEANS OF THE REDUCED BISERIAL COEFFICIENT OF CORRELATION AND REDUCED McCALL-LONG-BLISS VALIDITY VALUE

(1) Scores	(2) Total distribution of scores f_T	(3) Deviations d	(4) $f_T d$	(5) Scores of individuals failing item 1 f_{f_1}	(6) $f_{f_1} d$	(7) Scores of individuals failing item 2 f_{f_2}	(8) $f_{f_2} d$	(9) Scores of individuals failing item 3 f_{f_3}	(10) $f_{f_3} d$
100–105	2	16	32						
95– 99	3	15	45						
90– 94	5	14	70						
85– 89	10	13	130	..	..	3	39		
80– 84	10	12	120	..	..	6	72	4	48
75– 79	13	11	143	..	..	10	110	6	66
70– 74	15	10	150	..	..	8	80	10	100
65– 69	22	9	198	..	..	15	135	17	153
60– 64	28	8	204	..	..	22	176	16	128
55– 59	26	7	182	..	..	18	126	21	147
50– 54	18	6	108	..	..	16	96	9	54
45– 49	14	5	70	..	..	9	45	10	50
40– 44	13	4	52	2	8	7	28	10	40
35– 39	12	3	36	1	3	10	30	2	6
30– 34	4	2	8	3	6	4	8	1	1
25– 29	2	1	2	2	2	1	1	1	1
20– 24	1	0	0	1	0	1	0	...	0
Σ	198	..	1570	9	19	130	946	107	795

$k = {}^{1570}\!/_{198} = 7.93$ Item 1: $_rr_{bis} = (7.93 - {}^{19}\!/_{9})(0.475) = 2.76$

Item 2: $_rr_{bis} = (7.93 - {}^{946}\!/_{130})(1.787) = 1.16$

Item 3: $_rr_{bis} = (7.93 - {}^{795}\!/_{107})(1.360) = 0.680$

$_rV_{MLB} = 9(7.93) - 19 = 52.37$

$_rV_{MLB} = 130(7.93) - 946 = 84.90$

$_rV_{MLB} = 107(7.93) - 795 = 53.51$

6, 8, and 10 the frequencies of failing scores have been multiplied by the deviations of column 3.

Actual calculations are as follows:

(1) The frequencies of the total distribution of scores (column 2) are multiplied by the deviation values (column 3) and summated to yield $\Sigma f_T d(1570)$.

(2) $\Sigma f_T d(1570)$ is divided by $N_T(198)$. The resultant value of 7.93 is the constant which appears in the calculation of every validity coefficient.

For Item 1:

(3) The frequencies of the distribution of scores for individuals failing the item (column 5) are multiplied by the deviations (column 3) and summated to yield $\Sigma f_{f_1} d(19)$.

(4) Column 5 is summated to $N_{f_1}(9)$ and this sum is divided by $N_T(198)$ to obtain the proportion, q, of individuals failing the item (0.0455). $q/z(0.475)$ may be readily found by consulting Dunlap's table[4] or by dividing q by the appropriate value of z. The latter appears in tables of ordinates for the unit normal curve. The above values are entered into formula (D) to secure the validity value of 2.76.

A further practical reduction in the calculation of these item validities may be made by noting that McCall, Long, and Bliss[2,7,8] have recommended a validity value, designated by V_{MLB}, which is identical with the biserial coefficient except for the lack of z. It is written:

$$V_{MLB} = \frac{M_p - M_f}{\sigma_T}(pq) \qquad (E)$$

$$= \frac{M_T - M_f}{\sigma_T}(p) \qquad (F)$$

The effect of eliminating z from the biserial formula is to weight more heavily those items which approximate a fifty-per-cent difficulty. This is illustrated in Fig. 1 where are plotted the weights given the fraction $\frac{M_p - M_f}{\sigma_T}$ by the biserial coefficient (weight pq/z), *MLB* validity value (weight pq), and Fisher's t test of the significance of the difference between two means (weight $\sqrt{Npq}$).* The weights are expressed relative to the weight at fifty-per-cent difficulty. It will be seen that with a constant value of the ratio, $\frac{M_p - M_f}{\sigma_T}$, the *MLB* validity value will decrease much more rapidly than the biserial coeffi-

* Fisher's t test is given by: $t = \frac{M_1 - M_2}{\sigma_T}\sqrt{Npq}$. It has, as far as the author is aware, never been used as an item validity technique although there seems to be no reason why it could not be so employed. It is included in the chart to show the relationship between it and the other two constants.

cient as the per cent of individuals passing or failing the item diverges from fifty per cent.

Long, *et al.*,[7] have shown that the *MLB* value may be actually more desirable in certain situations. In comparing seventeen item validation methods, they conclude that when the various item-validity techniques are evaluated "on the basis of the tests resulting from their

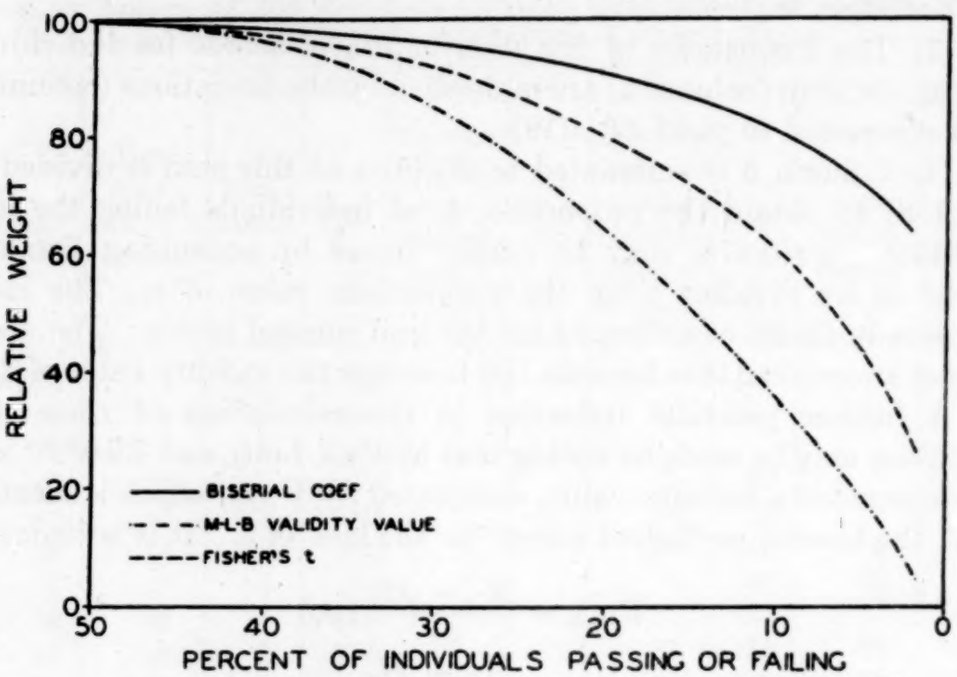

FIG. 1.—Showing the relative weights given the constant ratio $\frac{M_p - M_f}{\sigma_T}$ by the biserial coefficient of correlation, McCall-Long-Bliss validity value, and Fisher's *t* test of the significance of the difference between two means as the proportion of individuals passing or failing the item diverges from fifty per cent.

use, techniques such as the U-L Third, McCall-Long-Bliss, etc., which tend to discriminate in favor of items of fifty-per-cent difficulty, appear to be superior to techniques such as biserial *r*, Kelley, etc., which make no such discrimination." Presumably this would be true when the range of abilities tested is homogenous.[10,11] Other investigators who find that the best range of difficulties centers at fifty per cent are [3,6,9,13,14].

The application of our modifications to the *M-L-B* Validity Value with the additional elimination of an N_T yields an extremely simple formula:

$$_rV_{MLB} = N_f\left(\frac{\Sigma f_T d}{N_T}\right) - \Sigma f_f d \qquad (E)$$

The actual calculations for Item 1 are reduced to: Multiplying the constant value of $\frac{\Sigma f_T d}{N_T}$ (7.93) by the number of individuals failing the item (9) and subtracting $\Sigma f_{f_1} d$, the sum of the frequencies of individuals failing times the deviations (19). Thus:

$$_rV_{MLB_1} = 9(7.93) - 19 = 52.37.$$

A factor common to the two techniques discussed, and to Fisher's t, is the quantity, $\frac{M_p - M_f}{\sigma_T}$. Swineford[10] has found that the mean difference, $M_p - M_f$, is satisfactory as an item validity technique when it is not desired to weight the obtained validity for the per cent of individuals passing or failing. This would be the case when the range of abilities tested is heterogeneous. In such situations the mean difference should perhaps function better than the biserial r, which does weight the extreme ranges of difficulty a little (see Fig. 1). In Fig. 1 the weights for $M_p - M_f$ would coincide with the horizontal line at 100. It thus represents the other extreme of weighting from the *MLB* value.

Since the difference, $M_T - M_f$, is as adequate as $M_p - M_f$, the mean difference formula may be shortened to:

$$_rM_d = \frac{\Sigma f_T d}{N_T} - \frac{\Sigma f_f d}{N_f} \qquad (H)$$

The three reduced validity coefficients may be converted to conventional values by the equalities:

$$r_{bis} = (_r r_{bis}) \frac{i}{\sigma_T} \qquad (I)$$

$$V_{MLB} = (_r V_{MLB}) \frac{i}{N_T \sigma_T} \qquad (J)$$

$$r_{bis} = (V_{MLB}) \frac{1}{z} = (_r V_{MLB}) \frac{i}{z \sigma_T N_T} \qquad (K)$$

$$M_d = (_r M_d) \frac{i}{\sigma_T} * \qquad (L)$$

$$V_{MLB} = (M_d) \frac{N_f}{N_T} = (M_d)(q) \qquad (M)$$

$$r_{bis} = (M_d) \frac{N_f}{N_T z} = (M_d) \left(\frac{q}{z}\right) \qquad (N)$$

* For intertest comparisons the difference would have to be expressed in standard form by dividing by σ_T.

It should be emphasized that the shortened formulae proposed here yield values which are in every sense as accurate and adequate as the formulae from which they derive as long as the values obtained are applied to the particular problem for which they were computed. For the great majority of cases, this is all the teacher or clinical psychologist is interested in so that the simplicity of the reduced formulae should be quite helpful. If intertest comparisons are later thought necessary, however, formulae (*I*) through (*N*) will make the reduced coefficients comparable with no very great additional labor. As determined in actual practice, the time of calculation (excluding the tabulating of scores) may be reduced a third or a half.

REFERENCES

1. Barry, R. F.: "An Analysis of Some New Statistical Methods for Selecting Test Items." *Journal of Experimental Education*, Vol. VII, 1939, pp. 221–228.
2. Barthelmess, H. M.: *The Validity of Intelligence Test Elements.* Contributions to Education, No. 505. New York: Bureau of Publications, Teachers College, Columbia University, 1931.
3. Cleeton, G. U.: "Optimum Difficulty of Group Test Items." *Journal of Applied Psychology*, Vol. X, 1926, pp. 327–340.
4. Dunlap, J. W.: "Note on Computation of Bi-Serial Correlations in Item Evaluation." *Psychometrika*, Vol. I, No. 2, 1936, pp. 51–58.
5. Guilford, J. P.: *Psychometric Methods.* New York: McGraw-Hill Book Co., 1936.
6. Long, J. A.: "The Comparative Merits of Several Techniques for Determining Validities of Test Items." *Psychological Bulletin*, Vol. XXXI, 1934, 676, (Abstract).
7. Long, J. A., Sandiford, P., *et al.: The Validation of Test Items.* Toronto: University of Toronto Press, 1935.
8. McCall, W. A.: "Construction of the Multi-mental Scale." *Teachers College Record*, Vol. XXVII, 1926, pp. 394–415.
9. Richardson, M. W.: "The Relation Between the Difficulty and the Differential Validity of a Test." *Psychometrika*, Vol. I, No. 2, 1936, pp. 33–49.
10. Swineford, F.: "Validity of Test Items." *Journal of Educational Psychology*, Vol. XXVIII, 1936, pp. 68–78.
11. Symonds, P. M.: "Choice of Items for a Test on the Basis of Difficulty." *Journal of Educational Psychology*, Vol. XX, 1929, pp. 481–493.
12. Thorndike, E. L.: *The Measurement of Intelligence.* New York: Bureau of Publications, Teachers College, Columbia University, 1927.
13. Thurstone, T. G.: "The Difficulty of a Test and its Diagnostic Value." *Journal of Educational Psychology*, Vol. XXIII, 1932, pp. 335–343.
14. Uhrbrock, R. S.: "Analysis of 4,378 Test Items." *Psychological Bulletin*, Vol. XXXIII, 1936, p. 737, (Abstract).

THE DIFFICULTY REDUCTION TENDENCY IN PERCEPTION AND PROBLEM SOLVING

GERTRUDE HILDRETH

Teachers College, Columbia University

When children or adults are confronted with problems too difficult for them to solve, or situations beyond their understanding and experience, they tend to substitute simpler mental responses for the appropriate ones. For a required goal which is too difficult, an intermediate goal is set up. The natural tendency is to reach a solution that fits a simpler problem or situation, that allies the new problem to a familiar one, or that fits the person's insight, knowledge and experience. The problem is solved, not by straining to meet the difficulty, but rather by dropping it down to a simpler level. Piaget[1] observed that to understand certain notions and concepts the child has to deform and assimilate them according to a mental structure of his own.

The reduction process may be entirely automatic; ordinarily it is quite unconscious. A person's limited resources for dealing with the problem may prevent him from grasping the real problem in all its implications, and from seeing that his response is erroneous.

If people were less influenced by this principle in problem-solving they would more frequently respond in random fashion or give up all attempt at solution; but the fact is that random responses or "giving up" in ordinary daily life situations is infrequent as contrasted with simplifying the problems. The child in his normal daily activities, or the typical adult going about his routines, responding largely on a habit level, reveals in countless instances the operation of this principle, if solutions on different levels of insight and maturity are possible.

Psychologists are familiar with the principle of apperception early formulated by Herbert and elaborated by his followers. According to this principle, a person's mental responses are influenced by his experience and background. Herbert called this assimilative function of the mind "apperception." According to this doctrine, environmental influences determine which presentations the mind receives and also their manner of combining into higher mental processes. Acquisitions of new experiences are affected by those already acquired.

---

[1] Piaget, J.: *The Child's Conception of the World.* New York: Harcourt Brace, 1929.

Young children show the influence of apperception in many situations. One baby who was given some Phenolax asked for more "pink candy." The child who calls all animals "dogs" or all men "daddy" is responding in harmony with the "apperceptive mass."

The distinctive feature of the Difficulty Reduction Tendency as contrasted with apperception or "mind set" is the unconscious tendency to simplify the problem. This is not necessarily true in all responses affected by apperception or "mind set."

Errors and distortions are frequent when this tendency operates. These errors and distortions, though absurd, are not random nor chance. The response made to the simplified problem is a genuine, usually a meaningful, coherent solution. In fact, the tendency toward reducing the problem usually represents an effort to make the problem more meaningful. The movement is always in the direction from the less to the more meaningful.

Too frequently an individual's responses are rated as right or wrong with little attempt on the part of the rater to understand the significance of the particular response made. We obtain an excellent clue to the child's maturity in language development through studying his substitution errors in reading or speaking. The "boners" that cause teachers and parents endless amusement and that furnish the substance for our "humor" magazines are all first-hand data for the psychologist who would understand the operation of the Difficulty Reduction Principle. The errors that school children make provide analytic material for evaluating the child's mental development, understanding his thought processes and diagnosing teaching and learning situations. Through studying such errors we would be less prone to attribute errors to chance, or to be baffled by mistakes children make. Instead, we would have more sympathetic understanding of each child's learning difficulties.

These Difficulty Reduction Tendencies are shown in auditory and visual perception, in word meanings, in reasoning and thinking. The resulting errors are nowhere more commonly shown than in verbal responses due to the insecurity of auditory and visual perception. We say the individual misunderstood the verbal stimulus, that his hearing or vision was defective; the truer explanation is that he interpreted the question or problem in terms of his own mental grasp and insight which may have been inadequate for the problem. The chance that this type of error will appear in puzzle-solving or running a maze is less frequent.

Illustrations of the Difficulty Reduction Tendency are found in responses that involve reading, understanding spoken language, drawing or copying figures, and in broader thinking and reasoning problems.

ILLUSTRATIONS

Errors in Reading.—The person who makes an error while reading is seldom making a purely mechanical error. He is generally attempting to make sense out of what he is reading and in so doing he may perceive word symbols incorrectly. In his effort to make sense from symbols that are meaningless to him he may substitute his own preferred word symbols.

A schoolmaster in Virginia always called on one boy each morning to read a passage from the Scripture. One boy read: "Then shall the Kingdom of Heaven be likened unto ten Virginians . . . and five of them were wise and five were foolish." The schoolmaster stopped the boy and asked him to repeat the passage. He read exactly as before. "Well," commented the schoolmaster, "if the Bible says so, it must be true. But I never would have believed there were five foolish Virginians."

A young gifted child who had more zeal than ability made the following substitutions in her reading efforts, as noted by W. F. Dearborn:

"Ping" was read as "pink" or "pig"; "Forever" as "father"; "Organize" was read as "orange"; "Basket" for "breakfast"; "Tumble" for "theatre"; "Paints" then "Plants" for "parents"; "Business" as "Bess"; "Pleasant" as "plant"; "Friendship" as "friends." Another child read "large" for "lapse"; "purse" for "puree"; "ten fingers" for "ten figures." These were all easier substitutions for unfamiliar or difficult words.

A doctor's son read: "The children took off their stockings to wade," as "to get weighed." An adult read the book title "M Day" as "My Day," and was surprised not to see Mrs. Roosevelt's name on the title page.

Errors in Auditory Perception Shown in Vocabulary and Naming Errors.—It is a common tendency to interpret what we hear in terms of what we know or what we wish to hear. This makes what we hear easier to comprehend. We reduce auditory stimuli to our understanding level just as we tend to simplify visual percepts. The most frequent illustrations of these errors are found in the auditory perception of words and names. Many of our misunderstandings and

slips of the tongue fall under this heading. We say, "I didn't hear"; we mean, "I did not understand the words used."

Anyone who has asked children to respond with definitions to the words of the Stanford-Binet Vocabulary lists has frequently met this type of error, particularly in the young or dull child who quickly "gets beyond his depth." Since the words are given out of context, errors are inevitable.

A child of eight responded: "Guitar—you spread it on the streets"; "Muzzle—your arm is strong"; "Ochre—what you put in soup." "Roar" is confused with "raw" and "row," no matter how carefully it is pronounced.

A young child at church heard the clergyman read, "The *zeal* of thy house hath eaten me up," and asked his mother if he could see the animal after church. "My Country is a tree," one child sang lustily. Third-grade children reciting the Lord's Prayer in unison were heard to say, "Forgive us our trashbaskets and give them their trashbaskets."

High-school students made the following errors in themes.[1] The Crusades were fought in *Plasticine*. In a simple ceremony George Green, the president, handed over the *gravel* to the new encumbent. The fellows were cheered on to victory for our Alma *Martyr*. A new course prepares the student for the *vacations* for which he is fitted.

Children who labeled drawings of trains wrote most frequently on the coal car, "Cold" or "cool" car. A preacher reported that he was going to *officiate* elsewhere next Sunday afternoon, hence there would be no service. A child in the audience reported to his parents that the preacher was going fishing next Sunday afternoon and so there would be no Sunday School. One teacher repeatedly told the beginning first-grade class: "Sit erect, sit erect." The children reported at home that she said, "Cigarette, cigarette," to them repeatedly. Parents complained that the teacher did not enunciate distinctly enough.

At the Chicago World's Fair the artistically illiterate, eager to see the widely advertised and famous picture of Whistler's mother, asked to see the "Whistling Woman." Students in statistics class referred to the "medium" of a group rather than "median."

In 1818 Beethovan received a Broadwood piano from London, but a recent New York newspaper article referred to the instrument as a "Broadway." An article in the French press devoted to American

[1] Salisbury, R. and Leonard, J. P.: *Making sense; how to say what you mean and understand what you say.* New York: Scott Foresman, 1936.

customs observed that "After dinner some Americans play *three-robbers* bridge."

Word Etymology is illustrated in the following instances: Soubresault (French) from Latin "supre-saltus," "leap over" has become *Somer sault* and *Summer set.* We associate it with summertime and landing on the grass. "Breukelen" became "Brooklyn" to English-speaking people. Paul Dreiser became "Dresser"; "Izzy" Baline is now Irving Berlin; "Pumpkins" were originally "pompions"; "Kind" became Anglicized as "Kid."

Errors in Visual Perception and Drawing.—The Difficulty Reduction Tendency is illustrated when children and adults attempt to copy or draw from memory figures that either are meaningless to them, which are difficult to draw, or for which the time exposure is too short to enable them to study the figure sufficiently to reproduce it correctly. Children show to a marked extent a tendency to simplify intricacies in ambiguous figures by drawing substitute objects that are meaningful to them or easier to draw for any reason. A tendency toward integration and unification is shown in drawings of both adults and children, apparently in an effort to simplify the percept. Rhythm, symmetry and proportion may be introduced where it was originally lacking, because these tendencies help to organize and consequently to simplify the figures. The tendencies also represent more "primitive" and habituated responses. In children's drawings this tendency suggests a trend toward mental economy and least effort.

In drawing, young children tend to substitute a circle for anything suggesting roundness. The tendency is to "square" anything with angles. Apparently the square is simpler than the rectangle because all sides are equal. The rectangle is usually shortened. In copying a diamond, the child of four makes a rough triangle. A still younger child, instead of drawing a square or diamond as directed, will say, "I'll make a bed," or "It's a windmill," and draws a crude cross. A boy of three persistently made drawings of locomotives and trains of freight cars, but invariably made sausage shaped units instead of cylinders or rectangles. This tendency persisted in his drawings for over a year.

When young children before school age are asked to write their names, a task that is generally unpracticed and too difficult for them, they tend to scribble in imitation of adult cursive writing or essay a few large capital letters, such as they have seen on their blocks. In response to the request to write his name one child of three years and a

half made the cross he had made many times previously in response to other items in the Binet test.

ERRORS IN THINKING AND REASONING

The illustrations cited in the foregoing sections suggest what we may expect to find when we watch for the operation of the Difficulty Reduction Principle in problem-solving, thinking and reasoning. The typical response when the individual meets a problem beyond his powers of thinking or comprehension is to reinterpret the problem through simplifying it until it falls into familiar grooves. Laurence Shaffer found this tendency in children's responses to cartoons. The subjects could usually give a response to the cartoons shown them, but the response given by duller or younger subjects was on an inferior level, mere description rather than interpretation. Only the most mature and brightest subjects saw the full meaning in certain cartoons, *e.g.*, those having political significance, though others could interpret the pictures in general terms.

A four-year-old was asked to do the "Patience" test of the Binet Series, which involves putting two right-angled triangles together to make a rectangle. This child, after listening to the directions, said, "Our teacher taught us that. I can make the Trylon," and proceeded to make an isosceles triangle of the two pieces.

A child seven years, five months of age with an IQ of 100, when asked, "Who discovered the North Pole?," answered "Oh yes, who discovered America?" The examiner said again, "North Pole." He responded, "Oh yes, who discovered America? Why, Columbus."

A four-year-old at the beach was told: "It's low tide." "Who tied it?," he promptly asked. A school boy, viewing for the first time the statue of Lorenzo de Medici, exclaimed "Oh look at the mounted cop." A child telling of his experience in watching a carpenter for the first time, exclaimed to his mother: "The carpenter was busy taking curls off a board." A child in a garden after a heavy rain inquired: "Mother, why are the flowers leaking?"

A child observed his grandfather at morning and evening prayers and asked, "Mama, what does Grandfather say to the chair when he gets down on his legs?" A child who had been irregular in school attendance finally refused to go at all. She told her mother that the teacher had threatened to "drop her from the register." She interpreted this as the furnace register.

A high school boy heard about "major powers encircling Europe" and wished to know more about *Major Powers*, whether he resembled Major Bowes, etc. A student reading *Hamlet* reported: "Then Hamlet took the body of Polonius and cut it up in pieces." Seeing the shocked surprise from the teacher he said, "Right here it is. I'll read it to you." "He hath drawn apart the body he hath slain."

Professor Muse tells of the impatient surgeon who stuck his head out of the patient's room and called sharply to a passing nurse, "Bring me a probe,—quick." The nurse, who had just that day received her cap, answered, "There is no Probe (probationer) on this ward, doctor, won't I do?" A secretary advised her employer that all zeros must hereafter be omitted from Western Union messages to countries at war. When the employer asked where she got that idea she showed him the instructions: "Not accepting cipher in messages to countries at war."

Children's attempts to solve arithmetic problems too difficult for them show the "reduction" tendency. They tend to respond habitwise, using the processes that are most familiar to them in ways that have become habitual. This tendency gives the impression that the results were obtained by chance. Addition and multiplication processes are used more frequently than any others.

Illustration: Question: If two pencils cost five cents, how many can you buy for twenty cents? Answer: one hundred.

Illustration from Piaget: Question: How much is five times faster than fifty minutes? Answer: Forty-five. Question: How did you get it? Answer: I took ten and ten, then ten and ten, then added five.

INTERPRETATION AND CONCLUSIONS

The errors and mistakes cited in the foregoing sections are not random, mechanical or attributable to chance, but may be ascribed to the tendency to reduce a problem to simpler terms more or less unconsciously when it is too difficult in its original form. In this sense the answers given are not even "wrong," though in an absolute or a standard sense they are. The reductions tend to be meaningful transcriptions. Only as a last resort does the subject fail or give up. So long as he can respond meaningfully, he makes some response. The human mind abhors the meaningless. The responses to stimuli tend to be dynamic, not mechanical. The subject remakes the problem so that it will be a meaningful whole for him.

This reduction tendency is not wholly intellectual. It is also emotional and involves feeling tone as well. The reduction is a comfort-seeking device. The unpleasant situation is avoided. Blocked or thwarted efforts have little pleasure tone; quite the contrary, and that fact motivates the reduction.

Whether or not the illustrations given reflect clever or stupid errors cannot be determined directly from the nature of the response. This depends chiefly on who the individual is, and how adequate are his background and mental maturity for dealing with the problem. Most of the illustrations cited for children under seven were responses of normal or bright children to problems obviously too difficult for them. Bright as well as dull persons make these errors, but the more glaring illustrations occur with children responding to adult situations, and dull illiterate adults.

Do these illustrations suggest mental inertia on the part of the subjects, a tendency to respond with least effort, a retreat to an easier level? Unquestionably many of them do. Mental inertia appears to be as true a concept as physical inertia. Any given mental state of rest tends to persist. Habit binds people to lower level mental responses than they are actually capable of making. People respond in the way they have become habituated. Some of this response adults will not readily outgrow. The child has more chance for habit modification since he is in the unfinished evolutionary growth process and will be subjected to training for a period of years.

The Difficulty Reduction Principle operates universally. The phenomenon appears at all levels of mental maturity. It pervades varied types of response, materials and situations. It appears at all levels of perception and reasoning. It is particularly common in verbal responses.

These findings have significance for education. Teachers should pay more attention to these errors as clues to children's thought processes, degree of maturity, and background. Many Reduction errors are natural but immature responses that will be eradicated when the individual attains better mastery or understanding. The need for ascertaining the levels on which the individual who is instructed thinks and reasons is obvious. The principle illustrates the need for adopting a rational attitude toward the child's errors. Teachers commonly misunderstand this type of error. They caution students: "If you were careful and always pronounced syllables clearly and took pains to distinguish similar spellings that are easily confused, you wouldn't

make such mistakes." We should put the proper interpretation on pupils' errors. Calling the pupil's attention to the error or suggesting in general the need for more precision, better attention, will not necessarily take care of the problem.

The phenomenon indicates the futility of talking to people over their heads, particularly children, without sufficient background and explanation. It suggests the need for keeping requirements of reading, spelling, and arithmetic, as well as material in the content subjects, at the child's experience and maturity level. Children should not be given new concepts, words to master, problems to solve before having the necessary background and meaning. When presenting new concepts we should not take for granted that children will assimilate the idea without explanation and illustration.

SPEED *VERSUS* COMPREHENSION IN ELEMENTARY READING

LUCILE HUDSON BRIDGES[1]

The purpose of this research is to compare the effects on comparable groups of certain specific training for speed in reading with the effects of certain specific training for comprehension, checked by a control group which read daily with no specific instructions.

No claim is made that speed and comprehension can be entirely separated. The training was rather a case of emphasis on rapid reading as opposed to emphasis on careful understanding of what was read.

PROCEDURE

Subjects were pupils from grades 4*A*, 5*B*, 5*A*, 6*B*, and 6*A* of a public school. Intelligence distribution of the school is normal, and home backgrounds of pupils range from relief to professional families.

All pupils used were given the Pintner Non-Language Intelligence Test, and the Iowa Silent Reading Test, Elementary Battery. The Pintner test was used to avoid intelligence ratings weighted with reading. Half of the pupils were given form *A* and half form *B* of the Iowa test to avoid the effects of possible inequalities of the two forms. The following classes of pupils were eliminated: (1) Pupils with Pintner Mental Indices below 30; (2) pupils known to have serious uncorrected visual defects (accordng to school health records); (3) pupils fifteen years old or older. From the remaining children three groups were formed, equated for Mental Index, age, sex, and Iowa Comprehension score. One group was trained for speed, one for comprehension, and the third group was used as a control, reading in whatever way the pupils chose.

The texts used for training were Books IV, V, and VI of the *Friendly Hour Readers* by Leavell, Breckinridge, Browning, and Follis. Pupils were not permitted to use these books except at the prescribed reading periods. Children in grades 4*A* and 5*B* used Book IV, those in the higher grades, Book V. Children who finished one book were given the next book in the series.

---

[1] This study was made under the direction of Dr. J. B. Miner, Department of Psychology, University of Kentucky, Lexington, Kentucky.

TRAINING

The training in reading was given by the grade teachers. A twenty-minute period was set aside each day for forty school days during which the whole of each grade was engaged in some type of reading. The children were not aware that they were the subjects of a research since our method caused no departure from the customary school procedure. Each teacher gave in each group the following instructions:

Speed Group.—"The tests show that you need to read faster. Try to read as fast as you can. You will have a test at the end of each week to see how fast you read and how well you remember what you read, so do not skip any part of the stories you read. Do not read any part of the stories over again, as this will waste time. At the end of each reading period write your name and the date on one of these little slips. Write the number of the page you stopped on. Then count the lines from the top of the page down to the line you read last, and write the number of this line."

Control Group.—"Read in this book every day. At the end of each week you will have a test on what you have read."

Comprehension Group.—"The tests show that you need to read more carefully. You should think about what you read, so you will understand it better. You will have a sheet of questions to answer and things to find out in each story you read. Look at this sheet before you begin to read, and read to find the answers. Some of the questions you can answer as you go along. Others you may not be able to answer until you finish the story. Be sure that you get the right answer to each question, even if you have to go very slowly. You will have a test at the end of each week to see how well you have done this."

Teachers were instructed to give the children no help except to identify words not recognized, and to urge each group to concentrate on speed or comprehension, as the case might be.

After each seven or ten day period all groups were given a mimeographed test, not to obtain any measure of progress, but rather to make sure that the children did read the material and to let them think they were being checked.

The daily comprehension exercises were of the objective type, and had as their chief purpose the development of the following skills:

Noting facts and details, word meanings, summarizing, drawing inferences, and organization.

At the end of the forty working days the Iowa Silent Reading Test was given. Children who had form *A* at the beginning of the study were given form *B* at the end, and those who had form *B* initially were given form *A* finally.

RESULTS AND DISCUSSION

Table I shows the median Pintner Mental Index, chronological age, and number of boys and girls in each group.

TABLE I.—MENTAL INDEX, CHRONOLOGICAL AGE, AND NUMBER OF BOYS AND GIRLS IN EACH GROUP

Group	Mental index chronological age				Sex	
	Median	Q	Median	Q	Boys	Girls
Control	54	8.75	11-0	10 mo.	20	23
Comprehension	55	9.25	11-0	10 mo.	21	22
Speed	54.5	8.5	11-2	10 mo.	20	23

Inspection of Tables I and II reveals that the groups used were slightly above average in mental ability, due to the exclusion of children with mental indices below 30. The groups were slightly

TABLE II.—IOWA SILENT READING TEST: COMPREHENSION SCORES

	Control Group	Comprehension Group	Speed Group
Number of cases	43	43	43
Initial test:			
Median (points)	61	59	63
Q	26	23	23.5
PE of median	4.96	4.38	4.48
Grade equivalent of median	5-4	5-3	5-5
Final test:			
Median (points)	68	72	62
Q	27.98	16.55	28.34
PE of median	4.98	3.15	5.34
Grade equivalent of median	5-7	5-9	5-4
Gain points	+ 7	+13	− 1
Gain in grade placement	+ 3 months	+ 6 months	− 1 month

below the grade-placement norms for their ages, but about average in reading achievement for grade placement.

Table II shows the initial and final medians, Q's, and probable errors of the medians for each group in comprehension as measured by the Iowa Silent Reading Test.

Since the distributions of the reading scores are not normal, the statistical methods applicable to normal distributions are subject to wide errors. Therefore, critical ratios are not the final basis for evaluating the differences found. The grade norms of the Iowa Silent Reading Test have instead been used as a basis for comparing the improvement of the different groups.

COMPREHENSION GAINS

Since the initial tests were given the third week in February, and the final tests the second week in May, one would normally expect between two and three months' gain in reading to occur between tests.

Table II shows a median gain in comprehension for the Control Group of seven points, or the equivalent of three months. The Comprehension Group, however, shows a median gain of thirteen points, the equivalent of six months, or twice the growth that would normally be expected during this time. The Speed Group, shows a loss of one point, a median loss equivalent to half a month.

Table III gives the number of subjects in each group whose initial comprehension scores were above, below, or equal to the initial median of the Control Group, and the number in each group whose final scores were above or below this figure. Here, again, it is apparent that the Control Group has shown some gain, the Comprehension Group even more gain, and the Speed Group little if any gain.

TABLE III.—NUMBER OF SUBJECTS ABOVE AND BELOW CONTROL GROUP'S INITIAL IOWA COMPREHENSION MEDIAN BEFORE AND AFTER TRAINING

Group	Before training			After training			Number of increases
	+	−	=	+	−	=	
Control	21	21	1	25	18	0	4
Comprehension	22	21	0	28	15	0	6
Speed	22	21	0	23	20	0	1

Considering the evidence presented in Tables II and III, it seems likely that over-emphasis on speed not only does not improve com-

prehension in the group as a whole at this level, but may actually tend to retard the development of comprehension.

SPEED GAINS

Table IV shows the initial and final speed median of each group, with the Q and PE of the median in each case.

TABLE IV.—IOWA SILENT READING TEST: SPEED SCORES

	Control Group	Comprehension Group	Speed Group
Number of cases	43	43	43
Initial test:			
Median (points)	18	16	20
Q	3.5	6.5	4.5
PE of median	.67	1.24	.86
Grade equivalent of median	4-10	3-10	5-7
Final test:			
Median (points)	22	23	26
Q	6.5	7.5	6.25
PE of median	1.24	1.26	1.23
Grade equivalent of median	6-4	6-7	7-7
Grain in points	4	7	6
Gain in grade placement	1 yr., 4 mo.	2 yrs., 7 mo.	2 yrs.

Here it will be noted that the Comprehension Group shows the greatest gain, which is the equivalent of two years and seven months according to the test norms. The Speed Group shows the next greatest gain, the equivalent of two years. The Control Group shows the least gain, the equivalent of one year and four months. The variability of the speed scores on the Iowa Silent Reading Test is large, as this score is taken from a single subtest with a time limit of two minutes. It is likely that there was considerably more practice effect on this than on the comprehension scores taken from the longer battery of comprehension tests, which would account for the apparently phenomenal gain in the grade equivalent of these scores. Assuming this practice effect to be equal for all three groups, the relative gains of the groups are of primary importance. Certain it is that the group given training in comprehension gained at least as much in speed as the group given training for speed, while gaining considerably more in comprehension.

Table V, giving the numbers of subjects above, below, or equal to the Control Group's initial speed median before and after training, shows a larger number of subjects in the Control Group crossing the median line as a result of training, the next largest number in the Comprehension Group, and the smallest number in the Speed Group.

TABLE V—NUMBER OF SUBJECTS ABOVE AND BELOW CONTROL GROUP'S INITIAL IOWA SPEED MEDIAN BEFORE AND AFTER TRAINING

Group	Before training			After training			Number of increases
	+	−	=	+	−	=	
Control	19	21	3	30	11	2	11
Comprehension	18	24	1	27	14	2	9
Speed	25	17	1	33	9	1	8

A comparison of Tables IV and V shows that the Control Group had more subjects gaining in speed, but a lower median gain. This may indicate that reading at the child's normal speed is more beneficial to the development of speed than either pressure to read as fast as possible or the strong emphasis on comprehension resulting in slowing down the child's rate of reading during the training period. The data presented herein are far from conclusive on this point.

GAINS OF THE SUBGROUPS

The groups trained in fourth readers had initial Iowa Comprehension Medians as follows: Control, 4-2; Comprehension, 4-3; Speed, 4-3. Here the grade equivalents of gains were as follows: Control, seven months; Comprehension, nine months; Speed, two months. At this level, then, there is no question of the superiority of comprehension training over speed training for the development of comprehension. The poor showing of the Speed Group indicates that at this level forced speeding up of reading may inhibit the growth in comprehension that would otherwise take place.

In speed of reading, gains for these groups were: Control, two years, three months; Comprehension, one year, six months; Speed, three months. Whether or not the Control Group is really superior to the Comprehension Group here is open to question, in view of the wide variability of these speed scores. There is no doubt, however,

that both are superior to the Speed Group. At this level, then, either comprehension training or regular practice in reading with no especial emphasis is more effective than speed training for the development of both comprehension and speed.

The groups trained in the fifth reader are of a different character from those discussed above. They are composed of children slightly higher in Mental Index, slightly younger for their grade placement, and approximately two years ahead in achievement though only one year older than the lower subgroups. The Speed Group showed a gain of one year in comprehension, as compared to seven months for the Control Group and two months for the Comprehension Group. In Speed, likewise, the Speed Group showed the greatest gain and the Comprehension Group the least gain. Most of the children in these subgroups read very rapidly, and the Speed and Control Groups, without the slowing down caused by comprehension exercises, managed to cover more difficult material than did the Comprehension Group. Whether the differences in gains can be accounted for by actual emphasis on speed, amount of material read, or level of difficulty of the material read cannot be determined in this study. It may well be that at this level of development of the reading process, rapid reading becomes more beneficial to further development.

CONCLUSION

With those children in the fourth and fifth grades whose achievement was average or below, training that emphasized comprehension and took no account of speed was more effective in developing both speed and comprehension than was training that emphasized speed and minimized comprehension. Undirected daily practice in reading was also more effective at this level than training in rapid reading. As judged by these groups, over-emphasis on speed tended to inhibit growth in reading at this level. With the better readers (about sixth grade level) this did not seem to hold true.